

**Constructing City Images in Short Videos:
A Comparative Study of Beijing, Xi'an, and Shanghai**

Daqiang Li

A thesis submitted to fulfil requirements for
the degree of Doctor of Philosophy

Faculty of Arts and Social Sciences
The University of Sydney

2025

Statement of Original Authorship

This is to certify that the content of this thesis is my own work. This thesis has not been submitted for any other degree or purpose.

I certify that the intellectual content of this thesis is the product of my own work, and that all assistance received in preparing this thesis and all sources have been acknowledged.

No content produced by generative AI tools has been used in the preparation of this thesis.

Daqiang Li

August 2025

Abstract

This study examines how short videos facilitate the construction of city images through case studies of Beijing, Xi'an, and Shanghai, and aims to reveal both shared patterns and distinctive strategies in the three cities' image construction. Anchored in the city image communication theory and film aesthetics, it adopts the qualitative content analysis and comparative research, delving into the five representational techniques: color, light and shadow, sound, characters, and post-production.

The research findings show that creators' choices are shaped by a city's resources, development stage, and communication positioning, resulting in varied strategic emphases in each city's video cases. Beijing's short videos highlight solemnity and historical depth. Xi'an's short videos underscore traditional culture and everyday vitality. Shanghai's short videos blend modern dynamism with individual emotions. Meanwhile, as influenced by platform algorithms and the operational rationale of short videos, the five representational techniques operate in tandem. Color, light, and shadow jointly set the visual tonality. The sound system guides the delivery of emotions and rhythms. Characters serve as narrative anchors and experiential vectors. Post-production techniques organize and amplify the information expression, enhancing immersion and memorability of the viewing experience.

This study delivers academic contributions in three dimensions. Theoretically, it incorporates micro-level representational techniques into the digital communication framework, enriching the perspectives of city image research. Methodologically, it proposes a multi-dimensional analytical model tailored to the short video context. Practically, it offers actionable insights into future city branding and content creation. Overall, the study fully reveals how cities are constructed as both symbolic and experiential images in the contemporary Chinese media environment.

Keywords: city image construction, short videos, Xiaohongshu, Beijing, Xi'an, Shanghai, film aesthetics

Acknowledgements

I would like to extend my heartfelt gratitude to Professor Guo for his encouragement and insightful guidance, which have enabled me to grow both academically and personally throughout the completion of this thesis. I am also profoundly grateful to my second supervisor, Dr. Chen, for her constructive academic support and suggestions that significantly facilitated the research process.

Additionally, I would convey sincere thanks to my friends for their unwavering companionship and help along the way, and to my parents for their long-standing unconditional support. Finally, I wish to offer genuine appreciation to all those who have supported and assisted me in various ways.

Table of Contents

| | |
|---|-----|
| Statement of Original Authorship | i |
| Abstract | ii |
| Acknowledgements | iv |
| Table of Contents | v |
| Chapter 1 Introduction | 1 |
| Research aims and objectives | 1 |
| Research significance | 2 |
| Theoretical framework | 4 |
| Research methodology | 5 |
| Thesis structure | 9 |
| Chapter 2 Background of City Image Construction | 14 |
| Introduction | 14 |
| Historical background | 14 |
| <i>Demands of media transformation</i> | 15 |
| <i>Demands of urban development</i> | 17 |
| Literature review | 20 |
| <i>Studies on the constituent elements of Chinese city images</i> | 20 |
| <i>Studies on media strategies for city image construction</i> | 23 |
| <i>Research gaps</i> | 27 |
| Conclusion | 28 |
| Chapter 3 The Role of Color in Constructing City Images | 30 |
| Introduction | 30 |
| Color and natural landscape | 31 |
| <i>Color representation of seasons</i> | 32 |
| <i>Color representation of natural ecology</i> | 39 |
| Color and cultural landscape | 46 |
| <i>Color in historical and modern representation</i> | 48 |
| <i>Color in the expression of customs and traditions</i> | 55 |
| Emotional imagery of color | 63 |
| <i>Color tones and emotions</i> | 64 |
| <i>Color metaphors</i> | 72 |
| Conclusion | 80 |
| Chapter 4 The Role of Light and Shadow in Constructing City Images | 82 |
| Introduction | 82 |
| Natural light and shadow in city image construction | 82 |
| <i>Macro city aesthetics enabled by natural light and shadow</i> | 83 |
| <i>Refined aesthetics through natural light and shadow</i> | 90 |
| Atmospheric light and shadow in city image construction | 99 |
| <i>Spatial layering through atmospheric light and shadow</i> | 100 |
| <i>Atmospheric lighting and expression of city ambience</i> | 107 |
| Symbolic meanings of light and shadow in city image construction | 114 |

| | |
|--|-----|
| <i>Symbolic visual effects of urban light and shadow</i> | 114 |
| <i>Emotional symbolism of light and shadow</i> | 121 |
| Conclusion | 127 |
| Chapter 5 The Role of Sound in Constructing City Images | 130 |
| Introduction..... | 130 |
| Background music in city image construction..... | 130 |
| <i>Background music in emotional construction</i> | 131 |
| <i>Background music in cultural construction</i> | 139 |
| Sound effects in city image construction | 147 |
| <i>Sound effects in city authenticity representation</i> | 148 |
| <i>Symbolic meanings of specific sound effects</i> | 155 |
| Voice-over in city image construction | 162 |
| <i>Emotional expression through voice-over</i> | 163 |
| <i>Voice-over and narrative progression of short videos</i> | 170 |
| Conclusion | 176 |
| Chapter 6 The Role of Characters in Constructing City Images | 178 |
| Introduction..... | 178 |
| Driving the narrative forward | 178 |
| <i>Threading city scenes</i> | 179 |
| <i>Delving into city exploration</i> | 184 |
| Representing city life | 191 |
| <i>Depicting everyday city life</i> | 192 |
| <i>Conveying the city's spirit</i> | 197 |
| Evoking reflections on cities..... | 202 |
| <i>Reflections on the city</i> | 203 |
| <i>Reflections on individuals</i> | 209 |
| Conclusion | 214 |
| Chapter 7 The Role of Post-production in Constructing City Images | 217 |
| Introduction..... | 217 |
| Editing rhythm in city image construction..... | 217 |
| <i>Editing rhythm and sensory experience of city spaces</i> | 218 |
| <i>Editing rhythm and transformation of city scenes</i> | 225 |
| Shot combination in city image construction..... | 232 |
| <i>Shot combination and multi-layered city display</i> | 233 |
| <i>Shot combination and emotional connection with the city</i> | 239 |
| Visual special effects in city image construction | 245 |
| <i>Visual special effects in city imagery enhancement</i> | 246 |
| <i>Visual special effects and visual supplementation of city information</i> | 251 |
| Conclusion | 256 |
| Chapter 8 Conclusion | 258 |
| Key findings of the research | 258 |
| Contributions..... | 260 |
| <i>Theoretical contributions</i> | 261 |
| <i>Methodological contributions</i> | 262 |

| | |
|--------------------------------------|-----|
| <i>Practical contributions</i> | 263 |
| Limitations of the research..... | 264 |
| Concluding remarks | 265 |
| References | 267 |
| Appendix 1 Video List | 278 |

Chapter 1 Introduction

Today's rapidly advancing information age sees how the evolution of media forms and the diversification of communication channels are profoundly transforming humanity's means of accessing information and knowing the world (Kaul, 2012). To illustrate, compared to conventional media, short videos, as a new media form for disseminating and constructing city images, are concise, immediate, and highly interactive. This helps short videos to draw a wider audience, vividly present a city's cultural and lifestyle features through multi-dimensional immersive experiences (Wang, 2016), and in particular, function as a growingly essential bridge for communication between users and cities.

This study sources videos from Xiaohongshu, a famed Chinese social media platform, and delves into case studies of Beijing, Xi'an, and Shanghai. Its analyses of these three cities, each boasting distinctive historical and cultural characteristics and various levels of modernization, fully manifest the diversity and richness of exclusive contemporary Chinese city images constructed by short video creators with a film studies perspective. In short, it intends to cover the role and significance of short videos in such image construction.

Research aims and objectives

The core research question is as follows: What are the similarities and differences in the ways that creators construct images of Beijing, Xi'an, and Shanghai on short video platforms? Consequently, this study explores a range of representational techniques used by creators and strategic differences during city image construction, and specifically, works to achieve the following objectives.

First, assess the application of cinematic representational techniques, including the

use of color, light and shadow, sound, character design, and post-production, in short videos to express a city's spatial structure, cultural meaning, and emotional vibe, and thus to show their roles in constructing city images. These techniques empower visual presentation, crucially shaping the urban characteristics and evoking emotional resonance among viewers. Findings will reveal how short videos transform cities from physical places into symbolic and emotional images.

Second, make comparative analyses of city-specific strategies employed in the representative cases of Beijing, Xi'an, and Shanghai. Influenced by varying historical backgrounds, cultural resources, and stages of development, each city differs in its priorities and approach to short video creation. This research investigates how creators adapt their content choices and opt for differentiated techniques and emotional expressions as per the characteristics of each city. As a result, it uncovers the interaction between city features and media expression, offering new theoretical perspectives and empirical support for perceiving city image construction in contemporary China.

Research significance

Research significance is shown at theoretical, empirical, and practical levels.

Firstly, existing research primarily concerns conventional media like films and documentaries, while the role of short videos in city image construction has yet to be systematically explored.

This study is thus grounded in the city image communication theory and film aesthetics to analyze the usage of representational techniques for city image construction in short video creations, addressing theoretical gaps in relevant research and implying novel research perspectives.

Furthermore, through the aforementioned comparative analyses of the three cities, how short videos accommodate cities' historical and cultural contexts, spatial characteristics, and development orientations are discussed, leading to distinct expressions of urban images. This comparative perspective can lend constructive theoretical insights to future research on city image construction as well.

Secondly, this study reflects multifaceted empirical contributions. First, it facilitates the understanding of how city images and other forms of identity are constructed. By analyzing short video content in case studies, it reveals how cities leverage representational techniques for image construction, helping to deepen insights into the underlying rationale and provide references for a broader scope of studies, including those on national and cultural identity.

Next, it delves into the key part of short videos in city image construction. Short videos, with visuals at the core, outperform conventional media in disseminating city images, and integrate multiple representational techniques, including color, light and shadow, and sound, which creates more dynamic and layered expressions for city images.

Additionally, their flexible editing and narrative strategies enable the diverse presentation of urban characteristics. This study unveils how short videos reinforce city image construction through representational techniques and highlight their exclusive significance in communicating city culture and fostering a sense of city identity as well.

Thirdly, this study holds significant practical value, mainly for short video creators and city administrators. The creators can gain further insights into how to enhance the expressiveness of short videos through representational techniques and thus develop

concrete strategies for constructing city images. For city administrators, this study provides guidance on how to support such creative practices in real urban management, for instance by offering cultural and recreational activities that attract creators and audiences alike. These initiatives enrich the content resources for city image construction

Detailed analyses of city image construction in short videos also enable this study to theoretically, empirically, and practically guide the construction of China's city images in the global media landscape.

Theoretical framework

The research design of the dual theoretical framework blending the city image communication theory with film aesthetics supports the conceptual groundwork and analytical tools, probing into how short videos on Xiaohongshu construct images of Chinese cities.

At the macroscopic level, the theory helps understand processes of transforming cities into symbolic and emotional constructs using media. Motivated by the works of scholars such as Kevin Lynch (1964), Simon Anholt (2007), and He Guoping (2010), it highlights the importance of mediated representation in shaping both domestic and international audiences' urban perceptions, and the fact that a city's image, besides unveiling its physical space, gets actively constructed through cultural coding, narrative frameworks, and emotional resonance as well.

At the representational level, film aesthetics offers a crucial lens for analyzing short video creators' city image construction. Building on the work of scholars including Lacey N., Peng Jixiang, and Zhou Xian, this approach enables a close reading of visual and auditory elements, including color, light and shadow, sound, characters,

and post-production techniques. These five dimensions are central to how creators shape atmosphere, convey emotions, and construct city images through short videos. By applying film aesthetics to such a context, this study investigates the technical and emotional strategies behind the construction of city images.

In the complementary and mutually-reinforcing dual framework, the city image communication theory explores what constitutes a city image within the media context, while film aesthetics focuses on how representational techniques shape that particular image. This integrated dual framework conveys images of Beijing, Shanghai, and Xi'an from cultural, historical, and spatial dimensions using short video works. It offers further insights into the evolution of cinematic language in today's digital age to support mobile, algorithm-driven, and emotionally engaging formats of urban storytelling.

Research methodology

This study adopts a qualitative multi-case study design to explore how short videos construct city images on social media platforms. The multi-case approach is appropriate for capturing the complexity and diversity of urban representation, especially across cities with distinctive cultural and historical backgrounds. It enables the research to make in-depth comparisons among Beijing, Xi'an, and Shanghai, identifying both shared strategies and city-specific practices in the use of representational techniques.

The three cities chosen for case studies are based on the following considerations:

First, they reside at the exclusive crossroads of history and modernity. The ancient and modern Chinese capital, Beijing, balances the inheritance of traditional cultural essence with modernization advancements (Wang, 2011). Xi'an, evolving from its

historical identity as the Silk Road starting point to a contemporary hub in the Belt and Road Initiative (BRI), exemplifies the enduring evolution of Chinese civilization (Ng et al., 2016). Shanghai epitomizes transformative navigations of Chinese cities by integrating its colonial heritage and current role as an international financial center (Horesh, 2014). This interplay of history and modernity provides rich cultural imagery and narrative materials for content creation, thereby opening up space for the presentation of city images in short videos.

Second, they represent the diverse and distinctive nature of Chinese urban imagery. The Chinese capital, Beijing, upholds its lasting dignity as a national political and cultural center and a contemporary enterprising spirit. The millennia-old city of Xi'an well represents the profound heritage and historical continuity of Chinese civilization. The dynamic metropolis of Shanghai functions as an international urban pivot full of fashion and vitality. Such diversity thoroughly portrays China's multifaceted urban images and facilitates in-depth analyses of differentiated strategies and presentation approaches to constructing city images in short videos.

Lastly, regarding international exchange, they hold significant strategic importance. Beijing, as a pivotal city for global engagement, serves as a key platform for China's dialogue with the world. Xi'an, standing at the BRI core, bridges China with Eurasia. Shanghai, being a pioneer of reform and opening up (Xing, 2024), leads international interactions. The image construction of these cities as major gateways to China's openness, on short video platforms, is crucial to expanding the global impacts of Chinese urban centers.

A lifestyle-oriented social media platform, Xiaohongshu, is used as a source of videos for this study. Its abundant, highly active user base comprises young users from diverse regions and cultural backgrounds (Zhao, 2024) who are the primary audiences for city image communication and trendsetters in urban culture, tourism,

consumption, etc. Their proficient media literacy and acuity to new forms of cultural expression make the platform ideal for observing the construction journeys of city images within the contemporary media landscape.

To ensure data consistency and analytical clarity, this study focuses exclusively on Xiaohongshu rather than incorporating multiple platforms. Using a single platform avoids the potential complications caused by varying algorithmic mechanisms, user behaviors, and content styles across different platforms, which could lead to mixed or incomparable findings. Furthermore, Xiaohongshu operates as an integrated platform accessible to both domestic and international users, unlike platforms such as Douyin and TikTok that function as separate versions for different markets. This unified structure ensures greater coherence in data selection and enables the study to better capture how city images are represented, disseminated, and perceived within the context of global communication.

For data collection, short videos are selected as per their popularity and relevance to the theme of city image, including: (1) place-specific hashtags such as #Beijing, #Xi'an, or #Shanghai, as well as tags related to urban space, culture, or tourism; and (2) a minimum threshold of user engagement, defined as videos created by users with over 2,000 followers and receiving more than 100 likes. These criteria ensure that the selected content is both relevant to the topic and sufficiently visible and influential across the platform. As a result, it is more likely to showcase mainstream aesthetic preferences and shape public perceptions of city images with higher effectiveness.

The short video accounts involved include both non-official and official content producers on the platform. These producers encompass individual users, professional studios and municipal publicity accounts, all of whom collectively contribute to the construction of city images through various representational strategies.

Videos created between 2020 and 2025 are chosen for analysis to guarantee the content's timeliness. These analyzed videos exemplify up-to-date images of Chinese cities and reflect creative strategies of today's young content creators. This time frame clarifies a shift in the construction of city images amid the short-form social media age, signaling new heights in Xiaohongshu's growth and influence.

For case analysis within each analytical dimension, representative short videos were further selected from the dataset to enable in-depth interpretation. These cases also followed three refined principles: (1) thematic relevance—the video must explicitly involve the theme of city image; (2) city specificity—it should clearly reflect the cultural features or spatial atmospheres of Beijing, Xi'an, or Shanghai; and (3) user engagement—the creator should have more than 2,000 followers, and the video should have received over 100 likes. Videos created between 2020 and 2025 were included to ensure the timeliness and representativeness of the data. These criteria ensure that the selected materials are both visible and influential on the platform, reflecting mainstream aesthetic preferences and communication trends.

The study employs qualitative content analysis as the main analytical method. Each selected video is examined to identify how color, light and shadow, sound, characters, and post-production techniques are used to construct city images. The analysis proceeds in three stages: (1) identifying key representational strategies; (2) interpreting their aesthetic and emotional implications; and (3) conducting cross-case comparisons to reveal similarities and differences among the three cities. Through systematic analysis, the research ensures interpretive consistency and analytical rigor, laying the methodological foundation for the discussions developed in the following chapters.

Thesis structure

This thesis, devoted to the city image construction of Beijing, Xi'an, and Shanghai in short videos, is divided into eight main chapters, progressing from the research background to multidimensional analyses and comparisons. Each chapter, centering around the core research question, works to distinctly highlight how these three cities are represented similarly and differently.

Chapter 1 unfolds with the core research question, “What are the similarities and differences in the ways that creators construct images of Beijing, Xi'an, and Shanghai on short video platforms such as Xiaohongshu?” The question is followed by an overview of the thesis structure to develop an all-around research framework, and then by an introduction to research methodology, including the selection of short video sources and the city selection criteria. Finally, it covers the practical implications and academic contributions of this study.

Chapter 2 consists of the historical background and literature review. The first section studies the contemporary context of city image construction by investigating how the evolution of media ecology and shifting demands of urban communication have made short videos key to such construction. The second section summarizes existing literature on the components of city images and strategic approaches to media-relevant city image construction. The literature review identifies achievements and limitations of previous studies, thus setting up the theoretical foundation and abundant research space for this study and leading into the ensuing chapters.

Chapter 3 specifically examines diverse influences of color choices in short videos on city image construction. To begin with, it probes into the lens of natural landscape to elucidate how color represents seasonal changes and ecological characteristics. Next, its focus on the role of colors in the cultural landscape reveals how historical

elements, modern features, and local traditions are underlined. Finally, it analyzes the emotional imagery of color to find out the effects of warm and cool tones on audience perception and how specific colors metaphorize a city's cultural identity. Through assessing colors' application in constructing city images, this chapter demonstrates that color creates a distinct urban identity in short videos by simultaneously impacting visual perception and emotional engagement.

It is well-noted that color usage builds the emotional basis for the overall city image and subsequent applications of light and shadow, sound effects, characters, and post-production elements. It enables audiences to perceive city vibes visually, and lays the fundamental emotional tone and visual foundation for subsequent analyses. Besides positioning as independent visual research, such color analysis supports follow-up or future comparative studies in emotional and visual dimensions.

Chapter 4 is devoted to the influences of light and shadow on creating short videos with city image construction. First, it examines the role of natural light and shadow in both large-scale urban aesthetics and refined visual details. Next, it explores how atmospheric lighting augments spatial depth and cultivates particular emotional tones within the city. Finally, it inquires into the symbolic significance of light and shadow, revealing their differentiated occurrences in short videos of varied cities to communicate unique cultural identities and emotional bonds. In brief, this chapter demonstrates how light and shadow enrich city image construction by shaping spatial ambiance and underscoring temporal characteristics.

Light and shadow effects enrich the visual presentation of color, lending spatial depth and temporal flow to the integrated city image. Built upon the color-powered emotional basis, they construct concrete spatial experiences, supporting scenes and emotions for sound and character elements in subsequent chapters. Light and shadow also strengthen the emotional and spatial experiences in color and narrative,

significantly linking various chapters and deepening the three-dimensional urban impression acquired by the audience.

Chapter 5 studies the paramount role of sound in constructing city images via short videos. First, it examines how music cultivates the city's emotional ambience and conveys its cultural identity. Next, it investigates the role of sound effects in reinforcing a city's sense of realism or adding more profound symbolic significance through particular auditory elements. Finally, it analyzes the impact of voiceovers on the audience's perceived city images, particularly in emotional expression and narrative progression. This chapter highlights the utilization of sounds by short video creators to provide an immersive auditory experience, further consolidating the construction of city images.

Sound pushes cities' emotional expressions beyond the previously discussed emotional bases of color and light and shadow, so that the audience can comprehend the distinct atmosphere of city images in further depth. Incorporating sounds not only introduces an auditory perspective to the visual presentation of city images but also cultivates in the audience a stronger emotional resonance by enhancing the sentimental impacts of sounds with rhythm, melody, and timbre. As a result, sound, as the bridge for representational techniques for films and TV, links visual, emotional expressions and constructs more multilayered and cohesive city images.

Chapter 6 analyzes the role of characters in image construction from three key aspects. First, it centers on how characters propel narrations with their actions. Second, it explores the representation of urban life by characters through everyday experience and cultural spirit. Third, it examines how characters provoke reflections on the city. Character design emerges as a pivotal medium for emotional expression and conceptual projection, adding more intense humanistic meaning and real-world empathy to the city's visual representation, regardless of taking the city as a whole or

merely exploring the relationship between the individual and the city.

This chapter comes after the sound analysis and precedes discussions on post-production, building on the previous chapters' focus on sensory dimensions while shifting the priority to the distinctiveness of characters in constructing city images. As color and lighting shape visual styles, and sound reinforces the emotional atmosphere, the presence of characters brings vitality and a humanistic touch to the city's image. The rationale behind characters' actions and emotional portrayals also provides a narrative basis for post-production techniques. The research emphasis shifts from visual perception and emotional tone toward a more specific and affectively resonant representation of the city.

Chapter 7 explores the strategies employed in the use of post-production techniques. The first section examines how editing regulates the viewing experience through pacing alterations, reflecting the city's dynamic rhythm. The next section explores how shot combination enhances spatial depth and emotional presentation by varying camera angles and perspectives. Finally, the chapter investigates how special effects strengthen the visual appeal of city images and supplement their historical and cultural meanings. Together, these discussions reveal how the integration of post-production techniques and audiovisual elements contributes to the creation of more layered and coherent city images.

Post-production techniques refine and synthesize the representational elements discussed in previous chapters, including color, light and shadow, sound, and characters. They create cohesion among sensory experiences presented in short videos and enhance the expressive power of city images. By adjusting editing rhythm and applying special effects, post-production improves the emotional expressiveness of narratives and deepens their artistic impact. Through this seamless fusion of visual and auditory components, post-production plays a vital role in shaping city images

that are both coherent and emotionally engaging.

Chapter 8 sums up the main discoveries of the preceding analysis, including commonalities and differences in the representational strategies employed by creators across cities and interactive mechanisms among cities, platforms, and creators during city image construction. In light of these findings, this final chapter outlines the primary academic contributions of this study concerning theory, methodology, and practical applications, suggests research deficiencies in platform selection, and advocates for further endeavors of city image studies.

Chapter 2 Background of City Image Construction

Introduction

This chapter aims to outline the background of city image construction in China's short-video era in two sections, namely the historical background and literature review. The sections both expound the practical motivations behind such construction and offer theoretical support and academic positioning for this study.

The first section delves into the macro background. It begins with the media evolution and the transformation of city image communication strategies, and then explores the underlying reasons for the key mediating role of short videos in the context of city image construction. It argues that the application of short videos under this context not only actively accommodates media transformation trends but also responds to the changing demands of image communication and representational rationale in the course of city development.

The second section reviews existing research on the construction of Chinese city images. Its organization follows two major dimensions: studies on the constituent elements of city images, and discussions on the strategies for constructing city images in media settings. Based on the literature review, the section further identifies key limitations in current literature, particularly the lack of robust theoretical perspectives and methodological approaches, and research gaps in emerging communication mechanisms and creative practices on digital platforms.

Historical background

The rise of short videos as a major medium for constructing contemporary city images is not accidental but a result of the synergy of technological advancements and social

development. It is necessary to probe into the historical background of city image construction to truly understand this rising media form and its evolving representational rationale, such as by examining its media dynamics and underlying urban forces.

Demands of media transformation

The explosive growth of communication technologies has fueled significant changes in the media environment. Traditional formats such as text-based content and long-form videos are gradually being superseded by short videos boasting higher flexibility and efficiency. Short videos have both brought changes to people's daily means of accessing information and profoundly reshaped the pathways for constructing city images in the media transformation period.

City image construction aligns with the media usage habits and preferences of a new generation of users by adapting to media transformation patterns. As a result, it has elevated both the reach and emotional appeal of city representations (Jenkins & Green, 2013). The growing significance of short video platforms for the presentation of city images results from ongoing adaptations to the media transformation demands.

The core features of short videos lead to their natural alignment with the demands of city image dissemination. First, short videos present advantages such as a brief duration, fast pace, easy editing, and mobile accessibility, which endow them with strong competitiveness in today's media environment (Liu et al., 2022). Due to the highly fragmented attention span of contemporary users, cities are required to stand out from the abundant information flows by concentrated visual impacts and rapidly delivered information structures. This communication efficiency requirement is fulfilled by short videos.

Second, short video platforms support the algorithmic recommendation systems, which significantly enhance the precision and reach of disseminated information. According to Van Dijck, Poell, and De Waal (2018), short video platforms deliver content with user behavior data more accurately than traditional “linear” communication models. This has created favorable conditions for city images to reach target audiences with higher efficiency and align more closely with viewers’ interests.

More importantly, media transformation not just marks changes in communication tools but also signifies a deeper restructuring of the rationale behind representation. In response to this, the construction of city images gradually shifts from state-led models that emphasize authority and “grand narratives” toward narrative approaches centered on individual perspectives, emotional engagement, and living experiences (Li, 2023). The technical format and platform dynamics of short videos as a medium ideally enable this new form of image construction.

Creators can capture authentic scenes in daily urban life, such as street-level details, weather changes, cuisine culture, and emotional reactions with lightweight filming and editing techniques. These efforts help construct city images that evoke warmer and more relatable feelings.

Instead of constructing narratives with a single dominant voice, this decentralized approach to content production (Yang et al., 2024), upon integrating multiple voices, shapes city images that actively incorporate platforms and users.

This transformation also redefines who holds the power to represent the city. Before the rise of short videos, city image promotion was largely dominated by commercial and governmental capital, and urban representation was selectively shaped by economic priorities that projected an idealized vision of prosperity (Zukin, 2009; Harvey, 2002). The emergence of short video platforms has gradually disrupted this

capital-oriented pattern, granting individual creators new autonomy in shaping and circulating city narratives. Nevertheless, as participation expands, creators' personal motivations and performative tendencies may also blur the boundary between authenticity and fabrication, bringing certain negative implications for the credibility of city imagery.

In addition, the interactive and participatory nature of short videos further supports the augmented needs of city image construction. Online platform features such as likes, comments, shares, and challenge campaigns not only increase the frequency of interactions between audiences and content but also engage viewers in becoming participants and co-creators during information dissemination (Cao, 2021). This highly participatory communication mindset helps gain social traction around city-related topics, as shown in this study. It enhances the public visibility of city images and strengthens emotional engagement. The city, rather than a place of passive observation, emerges as a product for people to proactively experience and co-create.

In summary, concerning content structure, platform mechanisms, and dissemination models, short videos, as products of media transformation, are strongly consistent with the current needs of city image construction. Short videos hold prominently strengths in both form and function, such as concise information delivery, immediate emotional communication, and participatory platform interaction. Additionally, the city image construction has adapted to the shifts in media rationale, representing the advancements in communication strategy and a proactive adjustment in the approach to image construction. This process also sees the gradual popularity of short video platforms in shaping city images.

Demands of urban development

The rapidly progressing communication technologies have introduced significant changes in the media environment, leading to the continuous restructuring of social and cultural systems.

Cities, being a complex type of social organization, upgrade their development goals and communication requirements. Their modern development increasingly highlights media visibility and cultural output capabilities (Anholt, 2016). They have transitioned from passive recipients of national planning and promotional strategies to active shapers of cultural identity, constructing city images and engaging in international communication. Consequently, since short videos fulfill cities' multi-dimensional demands, they have developed into suitable tools for constructing city images.

Firstly, cities undertake growingly diverse communication functions encompassing tourism promotion, investment attraction, cultural exchange, and international engagement, among others. These all require higher flexibility of content and the diversity of formats (Holcomb, 2021). Conventional media, normally featuring linear logic and generalized messaging, struggle to meet the expressive needs of cities in varied communication scenarios and segmented audiences. In comparison, short videos are characterized by flexible formats, rich content, and a variety of platforms. This allows them to effortlessly fit into the city's image dissemination across communicative settings.

Secondly, city image construction shifts its purpose from satisfying purely functional goals to accentuating cultural meaning. A growing number of cities no longer rest on the simple target of higher recognizability and strive to cultivate a distinct and emotionally relatable "city personality" through nuanced visual expressions, localized emotional symbols, and perceptible cultural details (Kavaratzis & Hatch, 2013). The progressively saturated media context results in growing competition for user

attention, tourism development, and cultural influence among cities. Beyond seeking mere public visibility, they need to build emotional connections with the audience and shape their exclusive images.

Short videos offer favorable spaces to convey these “soft powers”. First-person narratives, sentimentally-driven editing rhythms, and everyday expressions all turn city images from abstract slogans to visual narratives that relate to viewers’ lived experiences. This shift from “information delivery” to “emotional resonance” makes city images more compelling and better harmonizes with next-generation audiences who are increasingly sensitive to authenticity, personal perspectives, and cultural distinctiveness.

In addition, the communication rationale of short videos is found to better match new forms of relationships arising between cities and target audiences. City communication used to follow the top-down information delivery pattern within platform-based ecosystems, but now transforms into a location to be observed, evaluated, and engaged with. Users employ features of likes, comments, and shares, and even recreate content on their own to be part of city narratives (Wang, 2023). This interactive model reshapes the mechanism of city image construction and dissemination.

Cities can precisely access potential audiences and alter their communication strategies based on ongoing interactions by leveraging platform algorithms and topic mechanisms, which enable more responsive and dynamic expressions. This user-participatory model of dissemination leads to a more mobile and open construction of city images, better catering to the fast-paced and fragmented features of today’s information environment.

Therefore, applying short videos to today's city image communication works as both a technical choice and a workable response to the evolving demands of modern city development. In detail, short videos communicate the essence of "what a city is" and "how to perceive a city" effectively. With the growingly emotional nature of communication and intense inter-city competition in this era, short videos as a whole have emerged as a key medium for constructing city images.

Literature review

This chapter has so far clarified the rationale behind short videos' key roles in image construction for cities, and will next examine the relevant academic research and theoretical assumptions. Though existing studies devoted to Chinese city image construction lend constructive theoretical references to this study, they hold theoretical gaps and analytical blind spots, which require further academic attention.

Studies on the constituent elements of Chinese city images

Existing research generally argues that city image construction is determined by the public's comprehensive perception and multi-factor value judgment, including urban space, culture, and social structure (Lynch, 1964; Chen, 2009). Studies on the composition of city images mainly concern the physical environment and the sociocultural dimension. Despite solid research progress in developing analytical frameworks, most studies remain focused on static categorization of components, and lack in-depth analyses of the mechanisms through which these elements are organized, transformed, and dynamically represented amidst city image construction.

As for the physical environment, architectural style and spatial layout are widely regarded as core elements of a city's visual identity. Landmark buildings often stand as highly recognizable symbols with strong cultural significance (Relph, 1976). For

example, the European-style historical architecture along the Bund in Shanghai epitomizes the city's historical trajectory and reinforces its image as a hub of "Haipai (Shanghai-styled) culture" and an "international metropolis" (Mei & Zhu, 2023). In addition, street patterns, public area distribution, and other structures of urban space crucially constitute the overall urban perception. The spatial impressions of "order" and "authority" are augmented by the central axis organization of Beijing, which is centered around the famed Tiananmen Square (Zhang,2024).

Although current studies underscore the visual recognition function of spatial forms, they pay less focus on how these structures interplay with emotions, cultural symbols, and ideology in mediated narratives to function as dynamic, symbolic forms of communication. For instance, the aforementioned central axis, though being a key symbol for national narratives, has yet to be fully examined regarding the means of transforming and culturally manipulating its meaning within the visual media context.

Infrastructure, as a symbol of modernity, has also risen as a research core of city image elements. In the case of Shenzhen City, its metro system and smart city initiatives are often considered living examples of "efficiency" and "technological advancement" (Chen et al., 2022). However, existing research mostly analyzes infrastructure from a functionalist perspective and thus overlooks its visual encoding and emotional appeals in visual media. For instance, in short-video creations, how transportation systems and smart facilities are transformed into trigger points of "modernity-related emotions" remains underexplored. The emblematic mechanisms of such representations within communication contexts have yet to be systematically examined.

At the sociocultural level, urban cultural heritage, traditional festivals, and local customs are extensively considered central to communicating the "spiritual image" of a city. Taking Xi'an as an instance, the city's visual reconstruction of "Tang Dynasty

memories” is facilitated by cultural revival projects such as the “Great Tang All Day Mall” (also known as the Grand Tang Dynasty Ever Bright City). In the case of Hangzhou, it shapes an image rich in Eastern aesthetics by promoting themes such as “Song Dynasty charms” and “Jiangnan poetic graces” at its communication core (Yao & Zhou, 2018).

These studies are oriented towards presenting and preserving cultural resources, but insufficiently consider the construction and reshaping of cultural content in media expressions. To illustrate, current studies on city communication are yet to fully address questions such as “Do content creators shape the narrative of Tang Dynasty memories?” or “Does this narrative involve emotional packaging and discursive reconstruction?”

Residents’ behavior and social atmosphere are also taken as important soft powers of the city’s image. Studies have shown that, viewing tourists and media narratives, public behavior and everyday interactions of dwellers are translated into perceivable urban impressions (Guo, 2024). Chengdu, to exemplify, is widely associated with a “livable city” image thanks to its “slow-paced lifestyle” and sense of friendliness. Nonetheless, the question of “How exactly is such an image constructed by visual and emotional cues in short videos?” remains understudied.

Furthermore, cultural and creative industries, as part of the broader social and cultural fabric, continue to receive scholarly interest. Cities enhance the recognizability and dissemination efficiency of their images by developing cultural brands and creative products. For instance, Chengdu leverages its iconic “panda culture” to facilitate its integration of culture and tourism, raising its domestic and international visibility (Wei, 2020). Nevertheless, despite this increased academic coverage and their dedication to product promotion and economic outcomes, most existing studies

initiate limited discussions of approaches to constructing city images, and hence are often confined to sales-oriented analysis.

How cultural and creative products frame a “simulated city” through audiovisual content on platforms, including short videos, is frequently investigated under the narrow framework of “cultural marketing”, which limits in-depth exploration of the representational functions.

In brief, current research on the components of Chinese city images has primarily adopted a tripartite analytical framework of “space–culture–society”, offering important insights into the question of “What a city possesses?”. The emergence of digital communication media, including short videos, witnesses the transformation of city image construction from traditional static display to a dynamic narrative model centered on visual language.

The majority of literature is threaded by mass communication and directs inadequate attention to the communication mechanisms and strategies employed by content creators of the emerging media landscape. As a result, they fall short of addressing the core issue of ways to construct city images on digital platforms. Guided by the existing research on constituent elements, this study will further delve into visual representation strategies and emotional expression pathways for short videos, exploring the reorganization and circulation of city images by such a new media form.

Studies on media strategies for city image construction

When propagating Chinese city images, various communication actors, expressive models of content, and media techniques collectively build up a complex mechanism of image construction. Relevant studies primarily cover two dimensions that both theoretically help understand the media-based pathways for constructing Chinese city

images: construction strategies of communication agents and the expressive methods of communication content.

In terms of agents for city image construction, research typically selects two types of representatives, namely Chinese government institutions and individual creators. Government-led city communication often adopts official channels and materials such as promotional videos, documentaries, or footage from large-scale events, all of which prioritize grand narratives, unified values, and visual order. This form of communication stresses economic development, cultural heritage, and social governance, using visual aesthetics and discursive disciplines to construct images of modernity, civility, and order (Yin & Gan, 2025).

Despite conveying the above priorities, the strategy often neglects engagements by the audience and the lived experience of locals, hence producing formulaic city images that are emotionally detached. In response to this, according to some scholars, “While propaganda-led city imagery may achieve communication efficiency, it often comes at the cost of individual perception and cultural diversity.” (Ke, 2024).

In stark contrast, following the emergence of social platforms like short video apps, ordinary creators grow into new agents for city communication. City images constructed by them more vividly present elements of everyday life, emotional empathy, and individual expressions in terms of both subject matter and representational style. They also probe into street-level documentation, citizen perspectives, and personal narratives (Qian, 2022). This approach departs from the overly rigid frameworks of traditional city communication, drawing city images closer to the concept of “city in everyday life.”

However, it is worth noting that some studies on individual creators remain constrained to conventional communication angles of platform behavior or user

preferences. They tend to ignore creators' strategic and aesthetic judgments when endowing urban meanings. Therefore, future research should take a deeper dive into how creators actively engage in the reproduction of city images through specific audiovisual techniques like framing, editing, and music selection.

Researchers often focus on the selection of visual symbols, the organization of narrative structures, and the communication of values. Landmarks, cultural venues, and festival events tend to be utilized as visual materials to construct a city's image, aiming to strengthen its recognizability and symbolic strength. This symbol-centered content strategy has also faced criticism despite its popularity. It inclines towards producing superficial, commercialized city images and is easily reduced to the state of reproducing "visual spectacles" and "stereotypical imagery". Additionally, a high degree of homogeneity is found in the symbolic elements selected for inter-city communication, heightening the difficulty of manifesting in-depth cultural structures and unique tones of each city.

In terms of narrative strategy, existing research highlights that city communication in short video creations values experiential and emotional considerations more. Creators construct domains of shared feelings through character-driven stories, fragmented scenes, and slices of time. For instance, they employ fast-paced editing to articulate urban dynamism, slow motion to emphasize historical vibes, and background music to arouse emotional resonance in the audience (Zhang, 2024).

Most of these studies still regard audiovisual language as a technical matter and insufficiently analyze its relationship with urban meaning through the lens of narrative mechanisms. Visuals and sounds, beyond being decorative elements to support communication, function as inseparable, narrative resources in city image construction. Obtaining insights into how they function in emotional expression,

rhythmic design, and urban perception continuously poses a challenge in today's research.

City image construction also operates as a vehicle for value communication. Drawing from existing research literature, concepts such as positivity, openness, diversity, and inclusiveness are often incorporated into city visuals through metaphors and symbolism, contributing to the projection of cultural identity and civic attitudes. For example, Shenzhen reinforces its image construction of a "futuristic city" through high-tech scenes, while Hangzhou adopts ecological imagery to underscore its advocacy of the "green development" philosophy (Peng & Du, 2024).

It is important to acknowledge that value construction in city imagery often follows an implicit pattern of disciplinary messaging. This is especially evident in government-led city communication campaigns where visuals are frequently used for the city's ideal portrayal while obscuring its real-world complexities and contradictions. Critical analyses of this prominent phenomenon are also inadequately addressed in studies. In particular, in-depth discussions of the underlying power structures in city image communication, including the message sender's and recipient's identity, are understudied.

Overall, regarding communication agents and content expression, current research on media-based strategies for constructing Chinese city images has established and applied a comparatively clear theoretical framework. However, eyeing the evolving digital media landscape, especially as short video platforms ascend to mainstream channels for any city's communication, existing studies are explicitly deficient in theoretical frameworks and analytical perspectives. Although some scholars have begun examining emotional expressions and visual guidance in short videos (Zhou, 2025), they haven't yet sufficiently assessed the mechanism through which creators contribute to shaping urban meanings with representational techniques.

Moreover, most studies adopt a single-city case study approach and lack cross-city comparisons regarding representational styles, image construction rationale, and cultural imagery within platform environments. Therefore, this study takes creators' strategies applied to short video platforms as a point of entry, delving into their construction practices through representational techniques. By conducting a comparative analysis of Beijing, Xi'an, and Shanghai, it works to further unveil the representation and reconstruction modes of city images in the new media era.

In recent years, research in urban studies and human geography has examined how digital platforms influence the representation and transformation of urban space. Researchers such as Butt (2016), Caprotti (2020), and Barns (2019) discuss the notion of platform urbanism, highlighting how digital infrastructures and algorithmic systems reshape the aesthetic and sensory experience of cities.

These studies reveal that platforms do not merely serve as channels of communication but actively participate in the production of spatial meanings and visual representations. However, much of this research remains focused on macro-level processes, such as the technological or infrastructural mediation of urban space. Few studies have investigated how such mediation is materialized through specific media practices, especially the creative strategies employed by short-video producers. This thesis builds upon these discussions by emphasizing the aesthetic and participatory dimensions of media practice in constructing city images on short-video platforms.

Research gaps

The literature reviewed above reveals two main directions. The first comprises general theories of city image communication, which conceptualize how urban identity is formed and disseminated through media and visual symbols. These studies

offer valuable conceptual insights but often treat city images as stable representational outcomes rather than as dynamic communicative processes. The second group of studies focuses on empirical research, emphasizing cultural promotion and media strategies. While these works provide abundant examples, they largely remain descriptive, lacking theoretical reflection on how city images are actually produced through media practices. Consequently, the theoretical and empirical dimensions remain insufficiently integrated, leading to a limited understanding of city image construction as an ongoing process.

This study addresses this disconnection by viewing city image construction as a media practice rather than a fixed outcome. It integrates theories of city image communication and film aesthetics to form a process-based perspective. The analysis focuses on how creators use representational techniques such as color, light and shadow, sound, characters, and post-production to shape meaning on short-video platforms. Through this approach, the study connects theoretical discussion with concrete media practices and explains how city images are composed, constructed, and disseminated in contemporary media environment.

Conclusion

This chapter has clarified the central issue of how short videos become a key medium in constructing city images by examining both the macro background and conducting a literature review. It highlights how short videos as dissemination tools are better positioned to meet the multifaceted requirements of city image construction in today's media arena through analyses of the parallel evolution of media technologies and ever-changing city communication demands. Additionally, it detects theoretical and methodological gaps, particularly in image composition, creative strategies, and cross-city comparisons, based on the systematic literature review of Chinese city image construction.

The ensuing chapters will build on the groundwork in this chapter, illustrate short video production strategies, and analyze the restructuring of urban resources by creators through visual content to construct city images. The study will also examine the commonalities and differences in image construction among Beijing, Xi'an, and Shanghai from a comparative perspective.

Chapter 3 The Role of Color in Constructing City Images

Introduction

Color, as a universal visual symbol, is instrumental in constructing city images. It performs as an intuitive sensory expression and a vital vehicle for conveying emotions and cultural meaning in the visually intensive medium of short videos. Exploring the representation of color in short videos and its cultural significance works to reveal its aesthetic value and cultural depth incarnated in city image construction, infusing refreshing insights into diverse expressions and rich meanings of modern cities.

The integration of color with natural landscapes, human landscapes, and emotional imagery supports a critical visual foundation for the inherently multi-dimensional construction of city images. Natural landscapes are characterized by colors that capture the rhythms of changing seasons and ecological diversity, and vibrantly represent the city's natural ecosystem. Human landscapes merge modernity with history, showcasing the cultural richness and complexity of urban life through color schemes. Emotional imagery imbues city images with deeper emotional empathy and exclusive spiritual essence by applying warm and cool tones and metaphorical expressions.

This chapter studies the shaping effect of color on the city images of Beijing, Xi'an, and Shanghai via short-form videos. The first section examines the color usage in natural landscapes, ranging from the portrayal of seasonal changes to the recreation of ecological features, unveiling that color invests urban natural beauty with a distinctive vitality. The second section delves into the employment of color in human landscapes, analyzing its contribution in connecting historical roots to modern trends and nurturing cultural symbols within the city. The third section investigates the manifestation of color in emotional imagery to how warm and cool tones and

metaphorical devices elucidate the emotional and cultural profundity of urban identity.

This chapter concludes by summarizing the shared features and distinctions in the three cities' color expressions and elaborates on how color facilitates constructing each city's unique image. In brief, it reveals how color functions as a visual language to exhibit prominent city characteristics through its application in image construction. This delivers precious insights into the emblematic significance of color in the course of contemporary urban development.

Color and natural landscape

Natural landscapes are essential to urban space due to their depictions of geographical features, ecological states, and changes of seasons within a city, and are vivaciously and crucially presented by colors in short-video works. Color, being an element with strong visual recognizability, enhances the expressiveness of natural landscapes and exerts a fundamental influence on the construction of city images (Cohen, 2001).

Natural landscapes are endowed with prominent visual features by the capturing and presentation of colors and become an integral part of visual identity shown in short video content.

Besides, seasons introduce distinct color characteristics to natural landscapes of a city and thus become central to fostering its natural imagery. The selection and demonstration of season-associated iconic colors by creators empower natural scenes in short videos with a strong sense of time and explicit visual traits (Lavrenova, 2023). Color not only mirrors natural states of each season but also visually strengthens the link between the city and time progression, lending a vivid temporal identity and visual distinguishability to the city image. The creator uses color to represent natural landscapes, producing a lasting visual impression of city images

(Fenwick, 2024), which supports the construction of a city image grounded in seasonal changes or temporal awareness.

The ecological elements of a city are also visually recreated on screen through colors by creators, delivering more expressive natural environments for the city (Hao & Ryan, 2023). The color options and combinations also make the ecological landscape more visually distinctive. In this sense, short videos highlight and enliven the city's natural characteristics by the visual reproduction of ecological colors, helping to construct an exclusive city image.

This section assesses the representation of seasonal and ecological aspects of the city and delves deeper into their beneficial effects for constructing city images.

Color representation of seasons

Seasonal shifts correspondingly bring distinct color variations to a city's natural landscape and develop visual components essential to its image construction (Spirn, 1988). In short video design, creators capture the signature colors of four seasons to unravel the dynamic transformation of natural scenery over time. This color application strategy forges stronger nature-city connections while simultaneously turning seasonal landscapes into symbolic imagery, which deepens the visual memory for the city image. The following section examines the color utilization in short videos for depicting the seasonal characteristics of Beijing, Xi'an, and Shanghai, and comparatively studying how color constructs the city image.

The short video "This Moment in Beijing is the 'Paradise on Earth' Once Praised by Lao She" unfolds a vigorous visual impression of autumn in Beijing through strategic use of color. Besides golden ginkgo biloba leaves and vermilion ancient architects that constitute the dominant color palette, warm tones run throughout the video,

creating a secure, gentle, and inviting atmosphere across the city space. In this case, color presents autumn in visual terms, fundamentally shaping the seasonal aesthetic in the city's image.

The distinct symbolic meaning is undertaken by the color combination of yellow and red, as seen from Fig. 1. The golden ginkgo biloba leaves reflect the rhythm of seasonal change and conjure up maturity and harvest. The vermilion tone, drawn from the visual tradition of Beijing's historical architecture, often arouses solemnity and elegance. This color juxtaposition creates a layered visual image that simultaneously points to both cultural heritage and natural rhythms, offering a unique sensory experience in the visual reconstruction of the city's image through the camera lens.

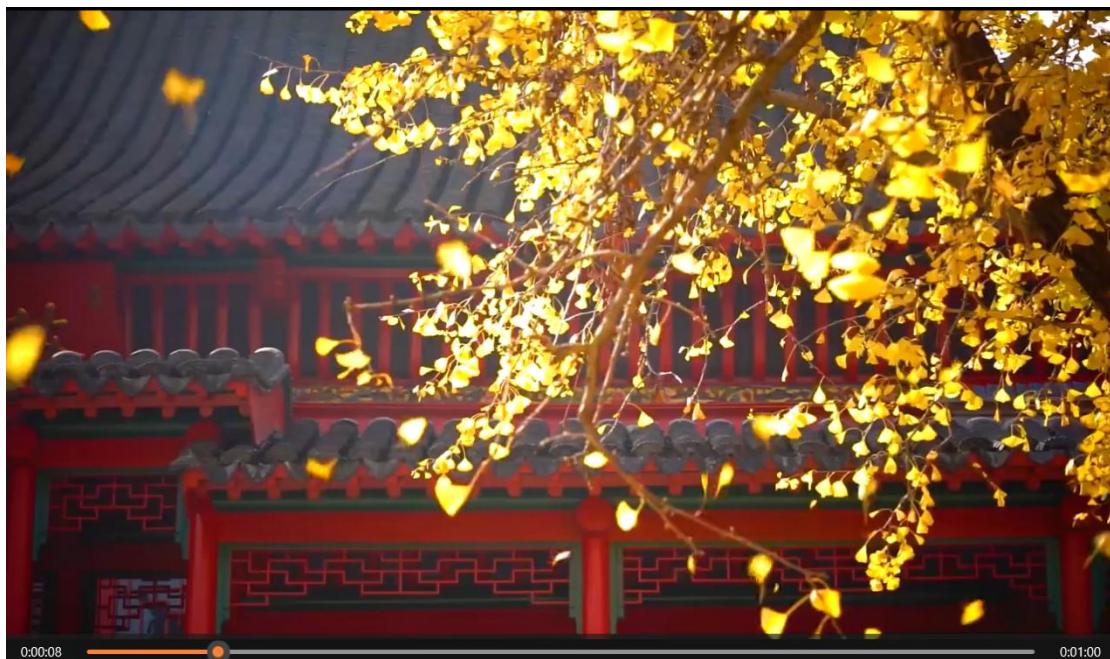


Fig. 1

The visuals are consistent with the unified color scheme (see Fig. 2) following the filming location changes throughout the short video. Golden and vermilion appear repeatedly as dominant tones, enabling color-based smooth transitions and connections across scenes. This approach positions color at the center of composition, stabilizing the visual rhythm arrangement and enabling viewers to grasp the seasonal

state of Beijing’s autumn spectacles at a glance. Colors also assist with organizing the landscape and perceptually guiding the viewers around the city.



Fig. 2

The creator explicitly conveys the typical elements of autumn in Beijing with colors. The interplay and contrast of warm-toned colors invigoratingly recreate the natural beauty and visual characteristics of the glorious season. Such a visual approach distinctly presents and communicates the autumn image of Beijing, and engages viewers to tour around the city’s bewitching natural landscapes.

The short video “May You Ride Swiftly with the Spring Breeze and See All the Blossoms of Chang’an in a Day” recreates the image of Xi’an in spring with rich colors. Sitting in a mid-latitude region with four distinctive seasons, Xi’an embraces the renewal season of spring. Its creator has filmed the imagery of blossoming flowers and picturesque natural scenery, and recorded the intensive engagements of visitors to solidify the city’s connection with spring.

The creator also emphasizes the saturation and layering of spring-specific colors. The viewers can watch how sunlight, surrounded by crystal-clear blue sky and pure white clouds above, and colorful flowers and greenery below (see Fig. 3), floods the scene. The color contrast makes the video more visually striking and turns the vitality of spring into the audience’s first impression of Xi’an. The creator employs this approach to sublime the concept of “spring in the city” into a dense color composition, thereby intensifying the sensory perception of Xi’an’s seasonal vibes.



Fig. 3

The presence of tourists further diversifies springtime color arrangements (see Fig. 4). As for the clothing styles of visitors, most of them wear bright spring colors, while a few still dress in darker winter attire. This contrast, preserved by creators, effectively highlights cross-seasonal transitions. This strategy that blends natural surroundings with human activity broadens the perception of spring from a period of environmental change to the exuberant everyday life. Tourists and landscapes together shape the overall impression of Xi'an in spring.

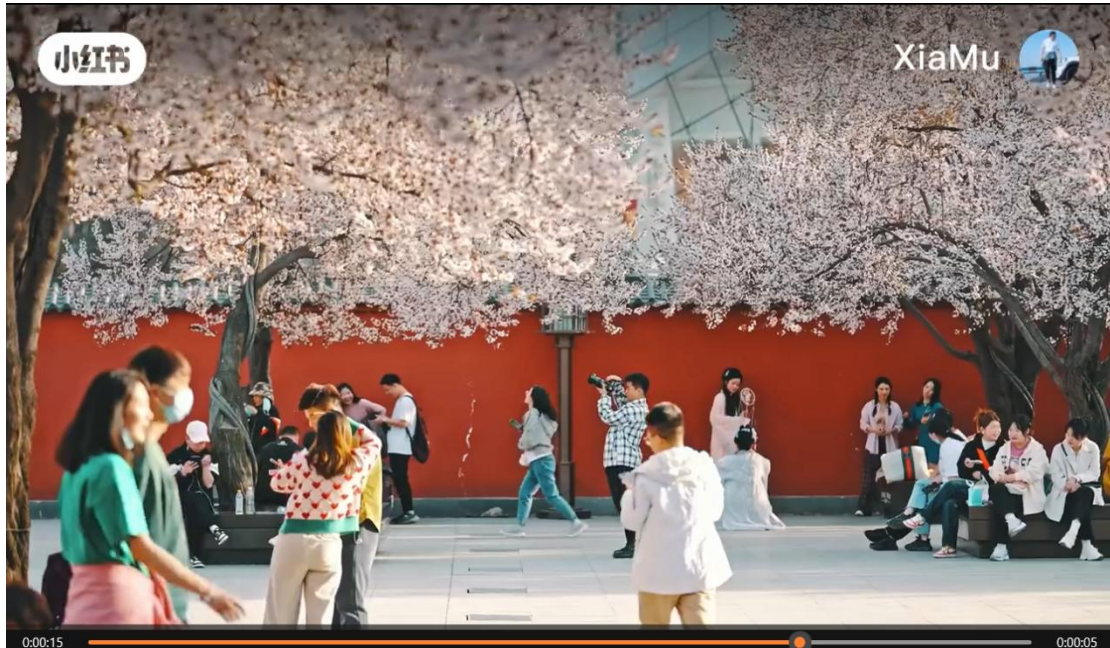


Fig. 4

The interplay between natural and human-related colors conveys a seasonal atmosphere distinctive to Xi'an. The creator captures subtle color differences from real-world scenes and intentionally combines colors to construct the city's image by closely adhering to the "spring" theme, thus creating a visual spectacle of the seasonal appeal of spring for viewers.

The short video "Perhaps Only Those in Sync Understand This Kind of Romance" seizes the colors of the plum rain season to reconstruct Shanghai's city image in this special period. The creator relies on color contrast of warm and cool tones, changes in light and shadow, and spatial relationships to reproduce the city's clear-cut climate features, and builds a uniquely ambient image of "Shanghai in summer" via visual expressions.

Contrasting between dark-toned interiors and lush greenery outside the window enhances the humid sensations in the plum rain season. The muted indoor colors bring about a tranquil post-rain ambiance, and vibrant green trees serve as visual accents, both evoking the invigorating and natural essence of the city after rainfall (see Fig. 5).

This color interaction not only articulates summer's environmental features but also adds visual depth to Shanghai's city image in this context. Low-saturation colors and softer light further underline the rainy season's moisture, using visuals to support perception of the seasonal atmosphere and reinforcing this particular experience in Shanghai's summer.



Fig. 5

Colors applied to outdoor scenes underline the post-rain visual characteristics of Shanghai. The city streets, pervaded by shades of green, feature tree canopies and climbing vines that highlight crisp air after the rainfall, further underscoring the humid summer atmosphere. While warm-toned buildings take up most of the frame, nuanced details such as green potted plants subtly balance the overall color composition, leaving a visually harmonious and pleasant impression (see Fig. 6). Such meticulous color arrangement in the short video partially offsets the sense of heatwave typical of summer, representing Shanghai's exclusive seasonal image.



Fig. 6

This short video integrates the precise use of color, coordinated application of warm and cool tones, and the interplay between people and their surroundings to visually recreate the characteristics of the plum rain season in Shanghai, making this period integral to constructing the city's image. The extensive use of color green amplifies the strong summer vibe and serene, atmospheric urban character of Shanghai by striking a visual balance between the post-rain dampness and summer heat in the plum rain season. The color-enabled depiction of natural landscapes works as a visual language to impart a tranquil emotional tone to the city, so that viewers immediately perceive Shanghai's exceptional traits in summer.

In conclusion, case studies of capturing and manifesting seasonal colors in Beijing, Xi'an, and Shanghai all stress their splendid natural landscapes, thereby strengthening the recognizability of their city images. The similarities and differences in color schemes represent how creators perceive the relationships between seasons and the city and adopt aesthetic choices.

Across the three cities, color demonstrates similar representational functions in depicting seasonal transformations. Rather than relying on single hues, creators adopt comprehensive color schemes to convey how natural cycles intertwine with each city's character. The combination of golden yellow and vermilion constructs the visual narrative of autumn in Beijing, evoking imperial associations rooted in its historical architecture. The multicolored palette of spring in Xi'an reflects the city's temperate monsoon climate and its lively, open atmosphere. The gradient from green to darker tones in Shanghai mirrors the city's plum rain season, embodying its humid serenity. This reflects how the use of color in all three cities embodies creators' ability to skillfully organize chromatic schemes in alignment with local characteristics, integrating tone coordination with regional distinctiveness.

In terms of color differences in videos, rich, unified warm-toned colors are applied to Beijing as representations of its stable, solemn image in autumn, lending cultural and natural richness to the city. The spring energy and vibrancy of Xi'an are outlined through saturation and a mix of warm and cool hues, which build an open, lively visual atmosphere. Higher tenderness and harmony are opted for in the color usage of Shanghai, presenting the fresh, calm imagery of the plum rain season with low-saturated greens and greys, and blending delicacy and refinement.

Despite the stylistic discrepancies, color plays a key part in seizing the natural glamour of changes in seasons, and constructs distinctive, recognizable city images in these three short videos.

Color representation of natural ecology

Ingenious selection and combination of colors help present and even strengthen the urban natural ecology in visual productions like short videos with enhanced visual expressiveness, investing the city image construction with fresh vigor, and making the

ecological environment central to shaping a city image. This section overall examines how colors enable short videos to reinforce the exclusiveness of city images in visual communication by accentuating the vibrant, characteristic natural ecological features, and contributing to the image construction of cities.

The short video “Beneath the Great Wall, Intangible Cultural Heritage Shines Through Time” unfolds the unique charm of Beijing’s natural ecology through thoughtful usage of colors, cultivating a stronger visual clarity and perceptibility in the city’s image. Colors are used to shape spatial relationships, highlight landscape features, and create contrasts that bring a distinct visual temperament to Beijing’s natural surroundings, developing a more recognizable image.

The organization of color is key to shaping the architectural composition. The green vegetation stands in contrast with the surrounding mountains, creating an open and well-defined spatial layout stressing the characteristics of natural scenery (see Fig. 7). Layers of green in varying shades illustrate the undulating mountain terrain, adding a sense of depth and stressing the ecological richness of the vegetation. Meanwhile, the grey-white walls run along the Great Wall.

This color contrast arrangement visually draws the viewers to gaze into the distance, creating a more tangible grandeur of the landscape. Additionally, it reinforces the spatial depth of the visuals while allowing visitors to swiftly identify Beijing’s natural environment, by prominently showcasing the ecological characteristics of the city.



Fig. 7

The creator employs cool-warm contrasts to increase the visual appeal of the natural landscapes in Beijing through color variations. Interweaving the cherry blossom red with green vegetation in the background produces natural color transitions that evade visual monotony but add vibrancy to the scene (see Fig. 8).



Fig. 8

Multidimensional color applications, including layering, spatial guidance, and smooth transitions, empower the creator to invigoratingly reproduce the ecological qualities of Beijing's landscape, infusing distinct ecological symbols and aesthetic expression into the city's imagery building.

The short video "A Thousand-Year-Old Meipi and the Snowy Qinling Mountains in One Stunning Frame", characterized by meticulous color arrangement and composition, unveils Xi'an's natural landscape in a more layered, inspiring manner. Techniques such as gradient colors, cold-warm contrasts, and spatial depth enable the creator to both forge a diverse natural scene and reinforce the bonds between the city and its surroundings, collectively constructing Xi'an's unique image.

Color gradation introduces spatial depth to the scene, heightening the perceptibility of the natural landscape layers. Camera shots record lush green trees featured in the foreground, a shimmering blue lake in the middle ground, and snow-capped mountains in the background. Such green-blue-white transition follows the natural order of spatial distribution and visually boosts the sense of vastness in the landscape (see Fig. 9). Viewers can explicitly experience the geographical diversity surrounding Xi'an.

The blue lake tone stands as a transition between the vegetation and snow-covered mountain peaks, keeping chromatic harmony within a cool color palette and prioritizing the significance of water in the city's natural context. This seamless color combination enables the visual coexistence of distinct geographic elements, reinforcing the image of Xi'an as a city abundant in natural resources.



Fig. 9

Color contrasts enhance the visual depth and tension, creating more vivid natural landscapes. Dark tones of the lakeside embankment distinguish water from the vegetation and visually guide viewers to gaze further into the scene. A warm-toned tall structure positioned at the lake center is distinct from the cool-toned background. This contrast strategy disrupts the visual monotony of dominant cool colors and emphasizes the dual presence of city-nature interplay and urban and ecological elements within the natural setting.

Through color gradients, warm and cool contrasts, and spatial design, this short video manages to record the grandeur and diversity in Xi'an's natural settings, thereby enhancing the recognizability of its city image in visual expression. The layered landscape presentation makes Xi'an's natural environment and city image more visually appealing to viewers, and color applications also boost the ecological character of Xi'an.

The short video "Half of Shanghai's Romance Lies on Wukang Road" stands for a careful orchestration of color composition and layered usage to capture the natural

charms of Wukang Road, illustrating the fascinating cityscape with a distinct visual tone. The creator deploys vehicles like color coordination, contrast, and visual harmony to blend urban space with natural elements, which shapes a gentle, approachable image of Shanghai. Color, as the visual core in the composition, is crucial to building the city's ambience and spatial imagery. The juxtaposition of nature and urban life offers viewers glances into the one-of-a-kind allure of Shanghai.

As for green elements that are pivotal for building urban atmosphere, lush trees lining the street and vine-covered building façades together create a symbiosis space of cityscape and nature (see Fig. 10). The unified tone of green raises the visual harmony and endows the city with sentimental tranquility and comfort, setting Wukang Road apart from other modern, fast-paced areas of Shanghai. This color arrangement signifies Shanghai's reverence and preservation of natural spaces when pursuing urban advancements, a testament to its city vision of human-nature coexistence.



Fig. 10

The color treatment in buildings is complementary to natural tones, enhancing visual cohesion in the city's image. Architectures in this historic district mostly feature soft,

warm hues that supplement the surrounding greenery, balancing the mix of warm- and cool-toned colors that enable a gentler and more welcoming urban environment. Tender colors of flowers also harmonize with the architectural tones, adding layers to urban space and using visual language to shape Wukang Road's warm, culturally rich atmosphere (see Fig. 11).



Fig. 11

Layered application of colors by the creator underlines and recreates the natural landscapes of Wukang Road as an urban space. Green tones demonstrate the tranquility and warmth in Shanghai, while brightly-colored flowers add an artistic touch to its surroundings. Meanwhile, the colors of buildings further enrich the overall urban ambiance. To summarize, color contributes to an amiable city image of Shanghai via visual expressions, supporting viewers with intuitive experiences of the city's unique charm of harmonious human-environment coexistence.

In all, the three short videos of Beijing, Xi'an, and Shanghai portray various color treatments that cultivate these cities' natural ecologies, accentuating stand-out styles and intricate connections between urban and natural landscapes. Their similar

approaches to color are accompanied by the distinguishable visual character of each city.

These videos all reflect the richness of natural ecological environments with color strategies to make colors integral to city image construction. The effective color usage enhances the recognizability of natural landscapes and the exclusivity of ecological environments in visual manifestations of short videos. The color combination and transition build layered compositions of vegetation, water, and mountains, highlighting the urban-nature integration. Techniques such as cool-warm contrast and light-shadow balance find their extensive applications in visual impact enhancement, fostering more straightforward and perceptible ecological features in cities.

The approaches to color expression vary significantly in the video cases of Beijing, Xi'an, and Shanghai. Beijing is inclined towards the steadiness and profundity of its natural landscapes, using color contrast and spatial composition to emphasize stability and vastness in its ecological scenes. Xi'an delves into the rich variety of the natural scenery, applying color gradations and changes to render a more dynamic, layered spatial effect. Shanghai places more emphasis on the city-nature harmony. In detail, it harmonizes natural scenes in the urban environment with the help of soft color transitions, building a more enjoyable, engaging city ambiance.

Regarding the image construction of natural ecology, color functions as a medium of visual communication and also vitally shapes the city's distinctive traits. Color expressions in the above three short videos depict seasonal changes and ecological features, and visually reinforce the cities' exclusive images to help viewers establish a rapid impression of the ecologies.

Color and cultural landscape

Cultural landmarks, as a vital part of city image construction, are ridden with historical memories, cultural identity, and elements of social life that collectively shape the symbolic character of a city. Color, a key visual communication tool, goes beyond representing physical places in short videos, and is essential to communicating the city's cultural depth and social atmosphere. Upon thoughtful selections and combinations of color palettes, cultural landmarks have their visual distinctiveness strengthened by creators, which quickly help viewers make first impressions on the cultural identity of a city (Chen et al., 2012). This process sees how creators translate cultural resources into visual forms that facilitate the city image construction.

Color bridges and balances the seemingly contrasting historical and modern elements. As historical sites often feature color schemes that represent the aesthetics and symbolism of that particular era, short videos are able to visually recreate the atmosphere of these historical spaces. Meanwhile, modern urban development introduces new color features. Short videos thus shape the visual dialogue between past and present, and a layered, multifaceted city image by showcasing the color diversity of these elements. Color usage not only strengthens the visual representation of historical culture but also underlines vigorous and diverse aspects of contemporary urban life, elevating the sense of time continuity in the city image.

Beyond expressing historical and modern charms, color also directly epitomizes local traditions and exclusiveness of everyday urban life in short videos, empowering the visual representation of local culture. Its usage enhances the visual appeal and recognizability of cultural activities, ranging from festivals, folk traditions, to daily street scenes, whose engaging nature is reinforced in short video production (Flueckiger, 2016). Color-driven recreations of cultural landscapes extend the city image from physical space to a richer structure that encompasses cultural ambiance and daily social experience.

This section presents the fusion of historical and modern elements by colors and examines its role in showcasing local customs and lifestyles, therefore highlighting its significance in shaping a city's cultural identity.

Color in historical and modern representation

In short video creations, color enhances visual appeal and serves as a meaningful vehicle for conveying the city's cultural identity and spatial characteristics (Wang, 2013). Both historical and modern components are amplified and vividly presented upon effectively capturing colors within urban spaces. Viewers can transform their phone screens into unique magnifiers for observing the city.

The following section investigates the case studies of Beijing, Shanghai, and Xi'an to help construct distinctive visual qualities in urban spaces by showcasing historical and modern elements.

In the short video "Life is Not Elsewhere", the creator naturally connects Beijing's ancient and modern identities with skillful usage of colors, unfolding a city holding on to its traditional identity while embracing the contemporary spirit. A variety of color elements, each boasting unique visual expressions, forge the synergy to construct the image of Beijing.

Eying the historical depiction of Beijing, red stands out as a core color amidst Beijing's cultural traditions. This color finds its way on red walls of the Forbidden City, doors and tiles of ancient alleyways, as well as in Beijing's local delicacies like candied hawthorns (see Fig. 12). Chinese culture takes the color red as the symbol of celebration, dignity, and heritage. The creator visualizes the time-honored character of Beijing as an ancient capital via the color red, which sublimes tradition into a cultural

symbol deep-rooted in the city's visual memory beyond the static architectural legacy. From towering city walls to neighborhood street snacks, this red palette conveys both historical gravity and the warmth of lifestyles in Beijing, epitomizing the diverse tradition and human bonding in traditional Beijing streets.



Fig. 12

When presenting modernity, the creator introduces a wide range of cool tones and vibrant colors to underscore the city's contemporary atmosphere and diverse character. Light blue skyscrapers emblemize technology and order and reflect Beijing's modern identity as a national-level political, economic, and innovation hub. Meanwhile, its avant-garde and varied lifestyles are displayed by the futuristic exhibition spaces and brightly colored ice cream displays (see Fig. 13).

Instead of stressing uniformity in these modern scenes, colors build up an all-encompassing context of openness, lightness, and exploration. This visual approach does not conflict with the red-denoted stability and solemnity, but complements it to diversify the city's color expressions, forging a dual emphasis on tradition adherence and continuous acceptance and development of novel ideas.



Fig. 13

The delicate color usage and combination help this short video capture the visual qualities of Beijing as a city fusing history and modernity. Despite a limited timeframe in the video, color manages to augment visual aesthetics and guide viewers to cast glimpses into the city's multi-layered space, thereby contributing to constructing Beijing's city image.

The short video "When Snow Falls on Chang'an, the Dream Fades Too" frames the snow-blanketed cityscape, seamlessly blends historical and modern elements of Xi'an. As enabled by the balanced integration of color uniformity, contrast, and highlights, the creator delivers a visual harmony between ancient city walls and the contemporary urban landscapes, constructing a city image of Xi'an where historical profundity and contemporary vibrancy coexist.

The creator sets expansiveness and serenity as the overall tone of the visuals by filming the snow-covered city. The blanket of white snow enables consistency in appearances of ancient walls and modern buildings, blurring the sense of time while creating a cohesive visual experience (see Fig. 14).

This low-saturation color treatment, besides underscoring the solidity and depth of historical structures, lends a subtle texture to modern architectures, displaying Xi'an's unique visual character that embodies both ancient and contemporary charms. The overcast sky further reduces brightness to enhance the sense of stillness and empower the historical allures of city walls to be distinguished from the monochromatic backdrop, reinforcing the sense of historical connection in the city's image.



Fig. 14

The incorporation of modern elements adds dynamic color variations to the scene, bringing liveliness to the otherwise static ancient setting (see Fig. 15). Visitors dressed in bright colors interact with one another when walking along the snow-coated city walls, emerging as visual highlights amid the low-saturation tones, and introducing a touch of vitality to the grand atmosphere.

Various colored cars pass through the historical gates, shaping an intense visual tension of the parallel timelines of present and past under the contrast between their vibrant tones and the grey-and-white background. These modern textures of color challenge the still impression of historical locations, showcasing the city's consistent vitality and modernity in the course of time. This interplay between movement and stillness registers a dynamic visual rhythm that nurtures a multi-dimensional, refreshing image of Xi'an.

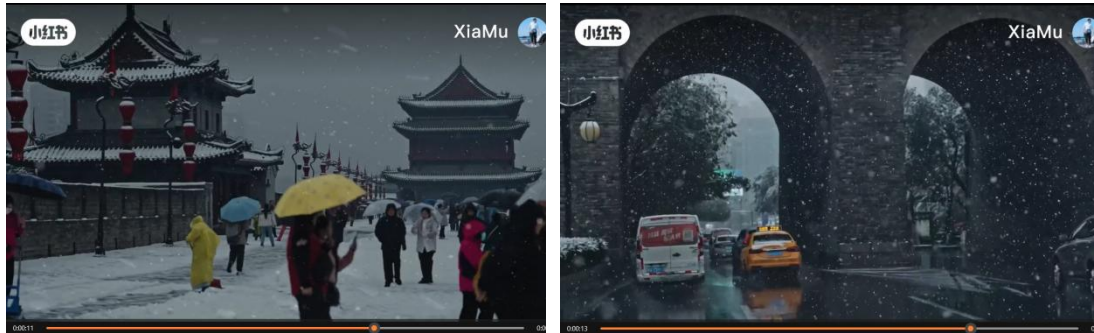


Fig. 15

Layered touches of color featured in the short video effectively display Xi'an's unique spatial identity at the intersection of history and modernity. Snow, as the base tone, visually unifies architectural elements across historical periods. The addition of modern colors brings a sense of movement. This collectively creates a city of deep-seated heritage and energetic vibes. This approach to color also satisfies multiple functions: reinforcing Xi'an's intrinsic depth as a city of historical and cultural significance; highlighting its modern appeals; and constructing the visual image connecting ancient with contemporary richness.

In the short video "Riding into the Summer Sunset, Watching Shanghai Dance a Tango", the creator achieves seamless connections between Shanghai's historical and modern glamour, portraying varied visual characteristics of the city across the entire day. By contrasting the colors of day and night, the video fluidly shifts between vintage and contemporary styles within the same urban space. In brief, it constructs Shanghai's image as an international metropolis that upholds the precious historical memory and disseminates constant flows of modern energy.

The daytime scenes, dominated by nostalgic tones, prominently narrate the city's historical traces and the continuity in its everyday life. Exotic plane trees in lush green, the red telephone booths, and green mailboxes work together to visually build a nostalgic impression (see Fig. 16). These colors, once written into the city's past, retain their endless charms in streets of today's Shanghai, symbolizing the long-

lasting historical traces in modern life. They reinforce Shanghai's image of fusing historical depth with daily friendliness, and add a touch of intimate, humanistic sentiment to the bustling and hustling urban lifestyles in modern days. This classic color strategy enables higher cohesion in the city's everyday scenes, stressing that Shanghai is an incarnation of the past and present by consistently depicting its ancient times and vibrant identity as a dynamic metropolitan city in the video.



Fig. 16

The nighttime scenes attach higher importance to colors of a modern metropolis, such as by applying neon lights and shifting light and shadow on building façades to bring out Shanghai's contemporary, internationalized character (see Fig. 17). Lights illuminating both sides of the Huangpu River cast reflections on water, and join the hustling traffic flows to create visual rhythm that presents the nighttime cityscape with a strong sense of motion. The integration of modern colors depicts the vibrant nightscape image of Shanghai, too.

The contrast between warm- and cool-toned neon lights, and the layered illumination of buildings, injects a vivid visual appeal to the nighttime of Shanghai, whose atmosphere differs from the daytime. As vintage tones of daylight evoke historical introspections, the modern glow cast by evening lights implies the city's future-oriented navigation. Shanghai's image becomes more diverse through this temporal color narrative.



Fig. 17

As empowered by alternating colors of day and night, the short video stands as a seamless integration of Shanghai's historical and modern vibes. Genial retro colors and the interplay of cool- and warm-toned modern lighting create a nostalgic and inspiring visual experience. Color, in addition to shaping the city's ambiance, connects the past with the future, presenting Shanghai as an inclusive, diverse, and multi-layered urban entity. Shanghai is thereby able to retain its unique cultural essence and undertake a more recognizable visual identity under its global metropolis context.

In conclusion, using colors to construct images for these three cities balances heritage with urban modernity and sustains the visual coherence. The contrast between warm and cool tones, along with saturation changes, allows traditional architecture, streetscapes, and modern spaces to coexist harmoniously within the same frame. Short videos simultaneously carry the profundity of historical culture and enhance the city's vitality through the integration of modern colors, enabling viewers to intuitively perceive the urban evolution through shifting color layers (Attademo, 2021).

In terms of differences, color-driven historical continuity, achieved with traditional colors extracted and then incorporated into modern design, infuses a sense of stability with the innovative short video of Beijing. The temporal overlap is highlighted in the short video presenting Xi'an, enabled by low-saturation tones to enhance the solemnity of historical architecture, while accents of modern color add visual dynamism. The time-based transition strategy adopted in the short video of Shanghai helps with the alternations between daytime nostalgia and nighttime neon lights. This brings forth a fluid, layered visual experience that reflects the historical and modern interplay.

The color schemes in the three cities seize the blend of history and modernity, and put stress on each city's unique character, hence elevating the visibility of city images constructed in short videos. This approach offers viewers an intuitive and visually dense immersion of the urban landscape.

Color in the expression of customs and traditions

In short video production, color, as a crucial visual language, is instrumental in representing a city's customs and local lifestyles (Liao, 2014). By capturing and employing colors, creators can vividly depict festivals, traditional practices, and the cultural atmosphere of everyday life, swiftly shaping a more distinct and emotionally engaging city image. The following section will dive into case studies of cities regarding means of applying colors to express local customs and cultural charm.

The short video "Winter at Shichahai: Don't Miss Beijing's Unique New Year Atmosphere" typifies the festive spirit of Beijing during the Chinese New Year through precise color usage, visually reinforcing the city's customs and cultural allures. The creator adopts skillful contrasts of warm and cool tones, which enable the stillness of winter to complement the vibrant festival colors for the New Year. This

contrast, by grasping the visual highlights of festive scenes, engages viewers to feel the city's cultural richness amidst traditional celebrations.

The scene, ridden with an overall grey-and-white tone typical of deep winter, creates a calm, steady urban atmosphere. This cool tone palette emphasizes the visual traits of Beijing in winter while adding depth to the spatial composition, which makes festive colors shine more brightly (see Fig. 18).

Against this backdrop, warm tones are introduced to bring rhythmic energy to the visuals. Red lanterns, Spring Festival couplets, and window decorations are conspicuously striking against the muted background, becoming visual focal points that accentuate the presence of New Year traditions within the cityscape. According to Yu (2023), creators can apply colors to enhance the emotional impact of a scene. In the camera shot with the window being pushed open, the yellow windowsill and red decorations instantly rise as visual focuses against the grey-and-white background. This color arrangement, as an effective visual language, allows viewers to quickly sense the festive ambiance and ritualistic sentiments of the Chinese New Year.



Fig.18

Following camera shots deep into Beijing hutongs, orderly arranged red decorations, as key visual markers that connect various scenes of the city's New Year traditions, appear across different spaces. From the red lanterns under the eaves to the Spring Festival couplets at the entrances of alleyways, colors function as ingenious visual guides. Apart from being a visual accent, the color red smoothly strings together citywide festive scenes by linking spatial elements through colors (see Fig. 19). Instead of the monotonous winter color palette, these warm-toned elements symbolize people's longing for reunion and lively celebrations during Spring Festival, offering a more tangible festive temperament of Beijing through visual expression.



Fig. 19

This color-applied short video successfully recreates the celebratory atmosphere of Beijing during Chinese New Year, infusing new life into conventional culture within the modern cityscape. The muted winter tones set the background, while the bright festive colors stand out as strong visual symbols, pointing to the holiday-specific cultural images of Beijing. Wu and Feng (2021) argue that color is capable of representing the particular cultural traits of a place. The creator's use of color leads

viewers into perceiving the festive atmosphere of Beijing in a short time. This, in turn, deepens their recognition of the city's fascinating character.

The short video "Chang'an, Ever Safe" employs striking color language to bring Xi'an's customs and city image to life through symbolism, visual contrast, and dynamic presentation. The colors applied recreate folk scenes and intensify the festal urban atmosphere and visual identity of the city by integrating color with space, people, and movement. Viewers can sense the city's cultural diversity and historical charm in no time by watching this video.

Red, the color running throughout the video, unifies the visual style and reinforces the symbolic significance of festive scenes (see Fig. 20). In the shots depicting traditional musical performances, performers in their uniform red costumes shape a strong visual cohesion. This color choice heightens the ritualistic feel of the scene and draws on the audience's cultural understanding of red, a symbol of warmth, celebration, and auspiciousness in Chinese tradition. Red becomes a key medium to communicate the festive atmosphere and folk culture, inviting viewers to quickly immerse themselves in a ritualistic festive context through this visual experience (Kim et al., 2023).



Fig. 20

The color-and-motion combination in this video enhances the emotional impact of the city's visual presentation. Dancers in vibrant costumes gracefully perform beneath the ancient city walls. As they move, the bold colors of their clothes shift, creating an intense visual flow (see Fig. 21). This dynamic color presentation symbolizes the continuity and evolution of traditional culture, bringing new vitality to the ancient urban place during festive moments. Combining colors with movements, the short video reconstructs the visual traits of Xi'an's traditional customs and expresses the city's living and evolving cultural heritage through the compelling visual language.



Fig. 21

The color presentation in the short video fully showcases Xi'an's folk culture while shaping a city image abundant in the convergence of history and modernity. Color functions not only as a basic compositional element but also as a key medium blending collective memory with cultural heritage and contemporary vitality. Consequently, Xi'an is portrayed as a city boasting multi-dimensional cultural appeal. This color-based storytelling technique guides viewers to further delve into local customs. In a broader context, it enhances the city's visual recognizability, enabling the coexistence and extensive propagation of Xi'an's cultural richness and modern vibrancy.

The short video "Promise Me, This Year You'll Spend Christmas in Shanghai" precisely selects colors to showcase Shanghai's exceptional urban style during the Western holiday season of Christmas. Red operates as the central visual element and is combined with color contrasts, diverse decorations, and thoughtful spatial design by the creator to heighten the festive mood while highlighting Shanghai's open, inclusive postures in global cultural exchanges. This deepens the immersive holiday feel and

helps Shanghai construct a more recognizable, globally connected city image, particularly in international celebrations.

Red dominates the visuals and is central to shaping the rich celebrative vibes, whose presence is found across scenes and items, including clothing, store decorations, and street ornaments (see Fig. 22). As a color typical of the warmth and enjoyment of Christmas, red immediately immerses viewers in the holiday mood. The effective contrast between red and other colors strengthens the festive feeling as well, making Christmas scenes in Shanghai more visually appealing and Shanghai's visual allure more prominent during celebrative times. For example, green Christmas trees and grey tones of winter both set off red more strikingly, creating more vivid and dynamic effects for holiday visuals.



Fig. 22

Beyond accentuating the main color tone of red, the short video also incorporates a variety of color elements for atmospheric enrichment into festivals. Minions and other cartoon figures featured in street decorations, and white and golden Christmas ornaments all serve as visual highlights that complement the dominant red palette and add playfulness and energy to the scene (see Fig. 23). These cartoon figures in bright colors boost Shanghai's approachable and more inclusive celebrative feel of Christmas. Combining bold colors with animated imagery challenges the typical red-and-green pairing, offering an engaging visual experience with more diversity. In this

sense, Shanghai's appeal as an international metropolis cultivates its more open and multicultural celebration of Christmas.



Fig. 23

The construction of Shanghai's city image in the video follows the patterns below: enhancing the festive ambiance with the color red; increasing the varied color combinations with visual richness; and integrating cartoon figures and decorations. This color scheme adds an interactive sense to urban spaces, intensifying the holiday spirit. Additionally, it successfully recreates Shanghai's liveliness during the Christmas season, cultivating a more vigorous, perceptible city image during cultural exchanges and global celebrations.

In conclusion, the utilization of colors in short videos portraying Beijing, Xi'an, and Shanghai aims to convey local customs and cultural atmosphere, shaping more characteristic city images in both festive and everyday contexts. Their visual styles differ, while color patterns present common traits.

In terms of similarities, color is used to highlight festive moods and traditional customs, injecting a stronger visual identity into cultural activities (Li, 2023). High-saturation colors create more lively festivals, while visual depth is enabled by contrasts between warm and cool tones, which further enliven cultural scenes. This reproduces the visual features of festive events and presents urban culture to viewers in an immediate and accessible way, thereby deepening their cognition of local customs.

As for differences, Beijing's color expressions emphasize ritual and tradition, as shown from the video, where festive colors often contrast with the city's mellow tones for the representation of its cultural depth. The dominantly dynamic visuals of the ancient capital of Xi'an combine brilliant colors with the historic backdrop to highlight the vibrancy and continuity of its folk traditions. The video portraying Shanghai blends modern urban atmosphere with festivity, presenting the open, inclusive city image in a global setting with enriched color combinations and commercial decorations.

The three-city color presentation reflects respective stand-out cultural traditions and reinforces each visual urban identity. Color, as a cultural symbol, renders more recognizable urban identities in short videos, evoking immediate relatability between viewers and each city's customs and atmosphere (Chen & Chu, 2022).

Emotional imagery of color

The use of color bears emotional and symbolic meanings to produce more expressive visual content. Color tone differences help create specific moods, guiding viewers to emotionally resonate with the city and shaping the means of imagery perception (Cakmak et al., 2024). The careful selection and adjustment of colors transcends

simple visual representation, attaching a deeper layer of meaning and enabling viewers to comprehend and relate themselves to the viewing experience.

Tone plays a crucial role in emotional expression within video imagery. Factors such as warmth or coolness, brightness or darkness, and saturation levels determine the overall mood of a scene and influence the viewers' emotional responses to the city (Ford, 1994). Short video creators delicately regulate tones for atmosphere building, including warm, calm, or invigorating vibes, all of which help shape the city's emotional identity and affect viewers' intuitive perception of the urban culture and atmosphere.

In addition, color also incorporates metaphorical meaning into the construction of city images. Beyond visuals, creators intend to convey more profound meanings with colors, demonstrating cultural richness, stages of development, or a city's unique character (Cao, 2025) in the short video. Such symbolic use of color leaves more scope for interpretation, in order that viewers can acquire further insights into the city by watching the video.

Apart from working as a representational tool for urban landscapes, color operates as a symbolic medium in shaping city identity. This section analyzes the role of tones in emotional expression and how color, as a metaphorical device, influences the construction of city images via short videos, thus highlighting the significance of color in constructing emotional impressions of the city.

Color tones and emotions

Short video creations take color tones as key to shaping the emotional context of a scene and to establishing the emotional foundation of a city's image. The thoughtful application of tones cultivates a richer emotional significance and thus a more layered

and expressive portrayal. The following section will concentrate on how tone is applied and expressed in emotional expressions through case studies of cities.

The short video “Cinematic Autumn in Beijing Is a Slice of Heaven” adopts a consistently warm tone for presenting visual uniqueness and emotional imagery of autumn in Beijing, seamlessly blending seasonal changes with the city’s identity. The creator, by turning tones into means of emotional expression, displays tranquility and vigor during autumn. As a result, Beijing is portrayed as a place imbued with warmth and emotional and historical depth, helping viewers forge deeper emotional bonds with the city through visual experience within the video.

Warm tones that dominate the holistic visual style are meticulously used to evoke shared memories of autumn coziness. Color saturation is enhanced in hallmark autumn scenes featuring maple leaves, sunsets, and golden foliage. The tonal consistency keeps the entire visuals soft and textured (see Fig. 24). This develops a sense of comfort under Beijing’s tender sunlight as the colors infuse the scene with warm and serene touches, and a hint of melancholy sentiment. In park scenes, the prominent layering of red maple leaves functions as a natural focal point and a symbolic element, suggesting that the city retains its vitality through color options even during seasonal shifts.



Fig. 24

The composition and tonal arrangement of the shots further reinforce the emotional expressions. In upward-facing shots, maple leaves and red architectural details echo

each other due to similar shades of red, which form the main visual focus. Grey tree trunks and rooftops in cooler tones constitute the subdued background, creating a subtle visual tension (see Fig. 25). The inclusion of warm tones softens the potential sense of distance often accompanying autumn's cooler palette, and therefore introduces a trace of emotional warmth to the scene. This color treatment transforms the city from a passive backdrop into a lived-in, emotionally resonant space. As a result, Beijing's autumn upholds its tranquility while continuing to voice vibrancy and cultural depth.

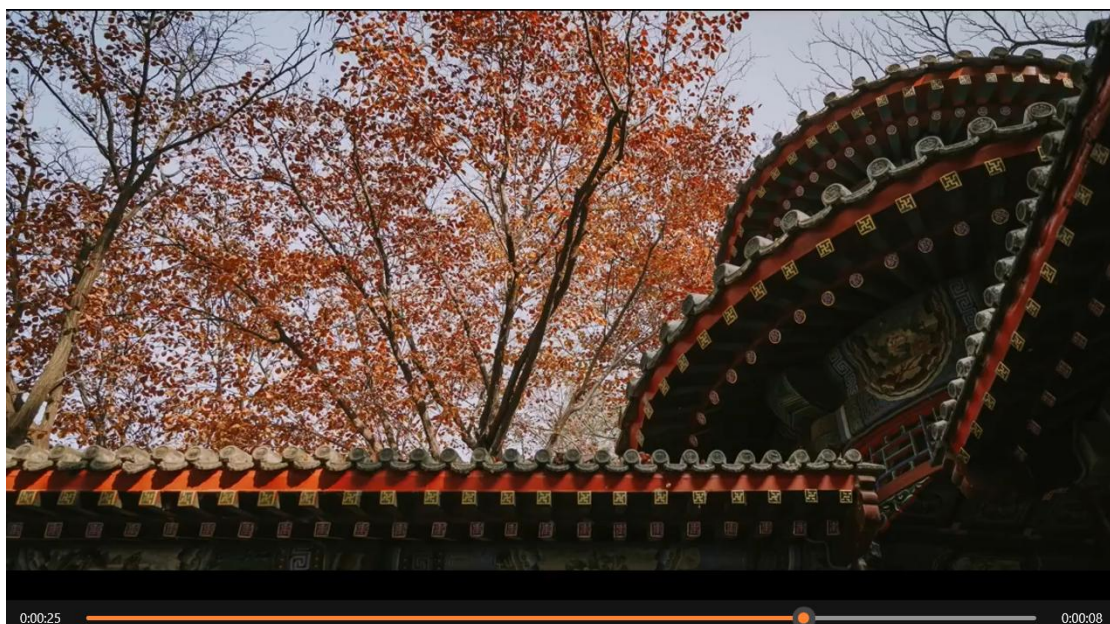


Fig. 25

Warm tones are applied throughout the video rather than being confined to single shots, maintaining a coherent visual rhythm and emotional thread. As the camera zooms between streets, buildings, and parks, viewers continuously perceive the warm tone as an emotional thread (see Fig. 26). This tonal consistency ensures smooth visual transitions and helps viewers coherently perceive the city's emotional atmosphere all across the shifting scenes. It also further reinforces Beijing's distinctive character in autumn. Repeated visual motifs, such as golden leaves and red maples, act as emotional markers running all through the video. They create a more

indelible visual impression of autumn scenery in Beijing, and make the city’s image more emotionally profound and recognizable.

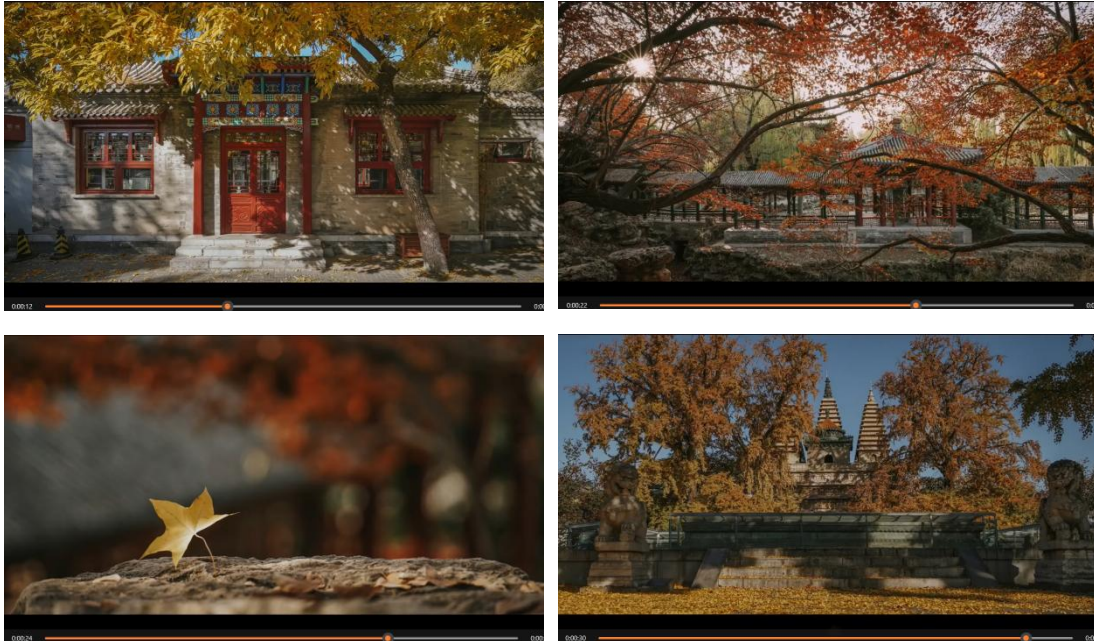


Fig. 26

The short video “Only Chang’an is Worthy of the Name of the Glorious Tang Dynasty” masterfully includes the cool-warm tonal interplay to present a rich nighttime landscape in Xi’an, while also merging historical depth with modern vitality in city image construction. The creator doesn’t just consider tones a visual styling tool, but also takes it as a key medium for arousing emotions and expressing the city’s spirit, closely bridging how Xi’an is perceived emotionally and visually.

In scenes around the “Great Tang All Day Mall”, warm tones serve as a central emotional guide. The combined use of orange and yellow lighting infuses night scenes with a soft, inviting glow, triggering associations with festivals, lanterns, and gatherings. These tones are widely available in streets, buildings, clothing, and people, which fill the frame with a strong lived-in sense and cultural warmth.

During traditional activities such as Hanfu (traditional Chinese clothing) performances, the costumes in vivid red and golden hues interact with the ambient lighting, further enhancing the festive atmosphere and cultural diversity that define the historical image of Xi'an (see Fig. 27). Warm hues, in addition to boasting visual appeal, conveys inclusiveness and sense of belonging, portraying Xi'an's nightlife as vibrant, open-minded, and rich in tradition.



Fig. 27

Upon shifting to a street performance beneath the city wall, the scene picks a cooler and more restrained (see Fig. 28) tone. The grey stone walls, dark-colored clothing, and low-saturation ambient lighting together build serenity with a touch of emotional distance. The cool tones depicted in this scene dually reflect the tranquil depth of historical contexts and shape a distinct emotional space for contemporary youth culture. Young people engross themselves in the stimulating rock performance under a millennium-old wall in the city, pursuing self-expression in a distinctly modern way. The subdued, introspective tone created by the cool tone palette offsets the tension between this modern representation and the ancient backdrop. Rather than

diminishing the city's warmth, it invites viewers to experience a more personal and inner-powered emotion.



Fig. 28

Employing emotional guidance and skilled tonal rhythms, Xi'an's city image evolves from a singular visual style to an authentic, poetic multidimensional narrative. The tone not just represents the city's external look but also unveils its deeper character and spirit, creating amiable and refreshing cultural charms at night for this ancient yet modern city. This approach, rooted in the relationship between tone and emotion, lends warmth and emotional depth to the city's image. It enables viewers to sense the energy and spirit possessed by Xi'an beyond appreciating the visual experience.

The short video "Plane Trees, Villas, and Alleyways—A Glimpse of Old Times" places warm tones at its visual core as well, capturing the relaxing atmosphere and leisure lifestyles in Shanghai by accurately applying colors and emotions. The maintenance of tonal continuity and alignment with emotional flows helps the creator seamlessly connect urban spaces with personal experience, and presents Shanghai as a city steeped in memory and warmth. Tone marks a stylistic choice as well as a

narrative thread that guides emotions and lends a gentle, nuanced expressiveness to the visuals.

The video starts with the soft yellow of old villas echoing the beige tones of clothing, whose color choice establishes an overall warm, soft palette (see Fig. 29), evoking a visual memory of the past while creating peaceful, friendly feelings. Regarding the interplay of light and shadow, shades of yellow and brown highlight the textures of historical architecture, naturally blending the elapse of time with everyday life. In this context, tone, besides being taken as a visual signature, serves as an emotional carrier that cultivates in viewers a sense of empathy and resonance through its splendid imagery. Viewers can also savor the tenderness and profundity rooted in Shanghai's city spaces.



Fig. 29

When zooming into more dynamic scenes, the coherent use of warm tones continues to present the city's emotional temperament. The dark red exterior of the double-decker bus holds on to the warm color scheme, bringing a sense of movement and modernity to the urban place (see Fig. 30). Red, whose richness and warmth guide the

rhythmic shift, transforms the city image from serene nostalgia to sentiments for the present. This does not break the emotional continuity, but instead, links historical memory to modern life and blends these two features naturally within the unified warm tones.

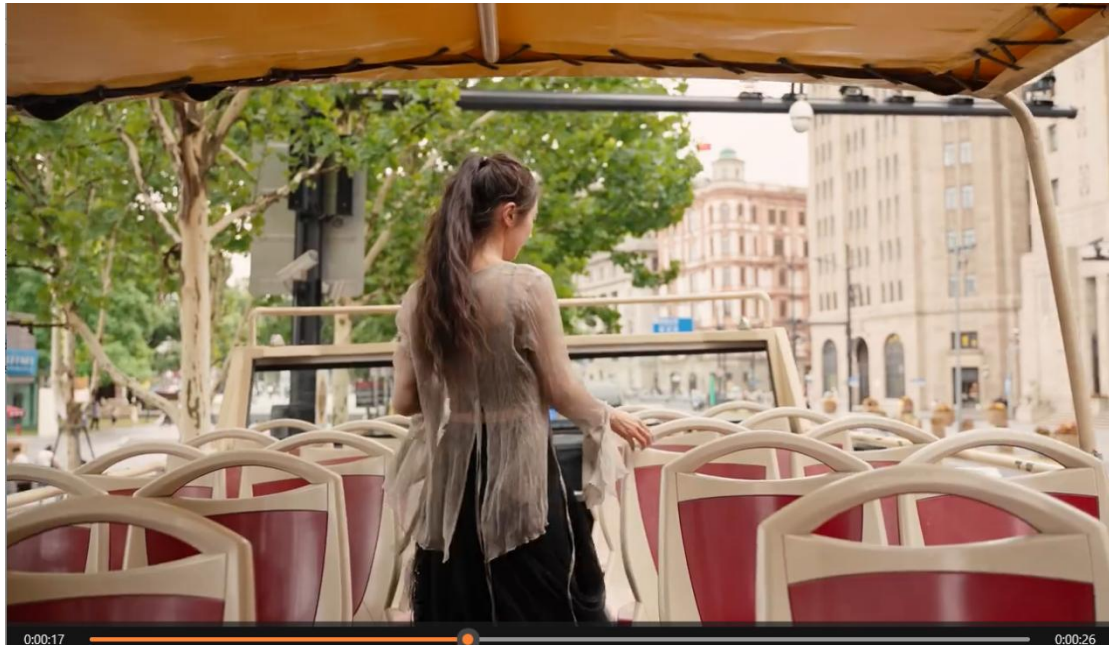


Fig. 30

Overall, the creator uses the intimacy between tone and emotion to depict a memory-abundant version of Shanghai whose warmth is deep-seated in humanity. This emotional strategy gradually shapes and reinforces the city's image using smooth, gentle, and tranquil visuals. This approach, where visual and emotional storytelling evolve in sync, cultivates a city image of Shanghai, which both holds strong visual appeal and a lasting sense of warmth and authenticity in viewers' eyes.

In conclusion, short video works feature distinct approaches to the interactions between tone and emotion in the cases of Beijing, Xi'an, and Shanghai, reflecting diverse strategies of emotional regulation in city image construction. All three cities share a creative rationale of tone-driven emotional atmosphere by amplifying the expression of urban culture and identity.

As for commonalities, tone is used for visual appeal, and more importantly, functions as an emotional narrative thread that navigates viewers into each city's emotional context. Warm tones are widely used to elicit warmth, intimacy, and peacefulness. Cool tones often find their applications in conveying tranquility, a sense of depth, or emotional distance. These chromatic schemes help audiences develop a quick initial impression of the city's character.

Emotional focus and tonal narrative trajectory differ among the three cities. The video depicting Beijing tends to consistently apply a warm palette to develop a city image replete with historical depth and warmth of humanity, emphasizing a peaceful mindset and emotional subtlety. The video portraying Xi'an, by contrast, shifts between warm and cool tones to represent its multilayered cultural identity, such as from vibrant and festive to restrained and introspective, or vice versa. This approach results in a more dynamic visual rhythm. A soft, warm tone is adopted throughout the video of Shanghai, integrating light and shadow with everyday details to express a relaxed lifestyle and warm nostalgia.

Overall, in spite of differences in the tonal strategies for the three cities in short videos, emotional depth is delivered through precise mastery of mood, enabling viewers to further inquire into each city's character through such excellent visual experience.

Color metaphors

Color as a visual language helps communicate the video creator's metaphorical intentions. In the video, the use of specific colors guides the audience into a multidimensional emotional space encompassing the city's external glamour and its in-depth cultural spirit. This organically builds a bridge between cultural transmission,

emotional expression, and visual storytelling (Xu, 2023) in terms of city image construction. The following section will investigate specific city cases to identify contributions by metaphorical color choices to the representation of both urban culture and individual emotion.

The short video “The Red Walls and Green Tiles of the Forbidden City: How Many Women’s Years Were Locked Away” connects Beijing’s historical depth with its contemporary cultural vitality using colors, building a rich and layered city image that bridges visual aesthetics and emotional resonance. The Forbidden City is presented as a cultural landmark whose iconic color palettes symbolize emotional and cultural continuity. This makes color a narrative tool that links tradition and modern components, empowering Beijing’s manifold and meaningful visual identity.

Red walls, golden tiles, and white snow of the Forbidden City as its dominant tones are featured in the first half of the video. These colors restore the architectural signatures of the royal complex and simultaneously hold meanings that symbolize the cultural and spiritual essence it represents. (see Fig. 31).

Chinese tradition often associates red with authority, dignity, and solemnity. Specifically, the use of “Forbidden City Red” highlights Beijing’s former status as an imperial capital and evokes collective memories tied to the city’s cultural roots. The scene of golden glazed tiles glowing under sunlight denotes glory and magnificence. Meanwhile, white snow covering roofs and ground accentuates the clarity and purity of the aforementioned dominant colors with its neutral cool tones, infusing the holistic composition with a historical and ritualistic sense. The cultural symbolic effect of colors set the core of Beijing’s city image: profundity, solemnity, and heritage.



Fig. 31

As the video approaches the end, the perspective shifts to how colors of the Forbidden City are recreated in contemporary life, showcasing paths where traditional color schemes transcend temporal boundaries to be ingeniously incorporated into modern urban culture. From cultural and creative products to everyday items, the creator keeps breathing new life into these color languages that embody ancient aesthetics by redesigning traditional hues such as red, gold, and blue (see Fig. 32).

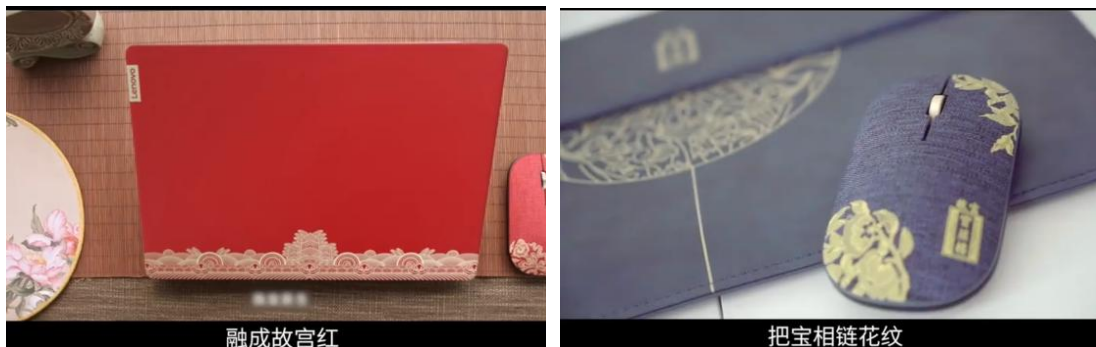


Fig. 32

This transition represents a continuation of visual style as well as an expression of cultural attitude, emblemizing how contemporary youth recognize and reinterpret traditional culture. Here, color metaphorizes consistent flows and the evolution of urban cultural spirits over time. As a result, Beijing takes on an image dually honoring history and embracing the future, in which its time-honored culture gets constantly preserved, refreshed, and reconstructed in a contemporary, trendy style.

Metaphorical color strategies help the creator elevate Beijing's city image construction from a superficially stylistic representation to a layered cultural narrative. And as these colors penetrate contemporary life, they reflect how tradition and modernity mingle and integrate with one another in the urban context. As seen from the video, Beijing embodies a particularly profound visual presence, not only as an icon of its imperial past, but also as a living synergy of cultural memory and contemporary vitality.

The short video "Snowy Day, Warring States Robes, and a Full Sense of Destiny" applies colors to a minimalist, refined visual composition to portray the profound, vibrantly tense cultural identity of Xi'an, a city brimming with historical heritage. By employing highly symbolic color elements such as white snow, red clothes, and a light yellow oil-paper umbrella against the ancient city wall as the backdrop, it visually unfurls Xi'an's millennia of history for the audience. This makes color crucially conducive to manifesting the city's spirit and cultural heritage. As the foundational visual language, color cultivates Xi'an's distinct urban character and emotional landscape through skilled contrasts between stillness and motion, coolness and warmth, tradition and modernity.

The video unveils with snow-blanketed city walls, setting a calm and solemn tone with a cool color palette that evokes a sense of historical weight and cultural depth. The snow white color symbolizes purity and quietude resulting from the passage of time, while also implying the lasting civilization of Xi'an, the once capital of thirteen imperial dynasties. The snow softens spatial boundaries to offer the city a timeless and almost eternal quality (Feng, 2019). Metaphorically, its whiteness sets a blank canvas background that is bracing for the arrival of colors. The creator introduces vivid hues of red and yellow to this background, creating strong visual and emotional contrast, and setting the stage for a powerful interplay between the city's historical stillness and the expressive presence of contemporary identity (see Fig. 33).



Fig. 33

Against the stark, wintry backdrop, a woman dressed in red emerges as the emotional and visual focal point. The light yellow oil-paper umbrella in her hand acts as a mediating color, delicately balancing the contrast between her red garment and snow whiteness. This soft hue of yellow, as a symbol of peaceful resilience, hope, and future considerations, outlines the warm tone of red, while conversing with the austerity of white. The umbrella itself suggests shelter and continuity, whose subtle hue, when paired with the figure's movement, visually presents the self-assured transition between tradition and modernity.

This color design thus shapes a more balanced composition. The creator exemplifies Xi'an's character in this video by using color strategies to display the city's adherence to traditions and courageous pursuit of modernity in ever-changing times.

By leveraging the symbolic and metaphorical functions of color, the creator substantially brings emotional and intellectual depth to the city's image. Cool tones in general hold the weight of history and personal memory. In warm tone expressions,

red conveys inspiring spirit and cultural lineage, while soft yellow denotes hope and the future. The interplay of these elements within the spatial composition and narrative flow together sketches a portrait of Xi'an, a city brimful with prolonged historical heritage and alive with refreshing expressions.

As observed from the video, the city anchors itself in cultural legacy while proactively seeking new strategies to keep up with modern trends. This minimalist yet profound use of color endows Xi'an with a dynamic and peaceful character abundant in profundity, diversity, and mobility. A condensed yet compelling cultural strength is articulated through visual language in this short video.

The short video "I Used a Cinematic Lens to Capture the Old Days of Shanghai" seamlessly blends the creator's in-person emotional experience with the fascinating cultural character of Shanghai, with color as a strong metaphor.

The overall use of dark tones set a nostalgic and quiet atmosphere (see Fig. 34), a tonal approach intentionally employed by the creator to stir visual memories of "Traditional Shanghai". This innovatively turns the cityscape into an extension of history. The red car, as an emotional vehicle, points to how modern people reinterpret the city's cultural heritage and spatial memory with their individual experience.

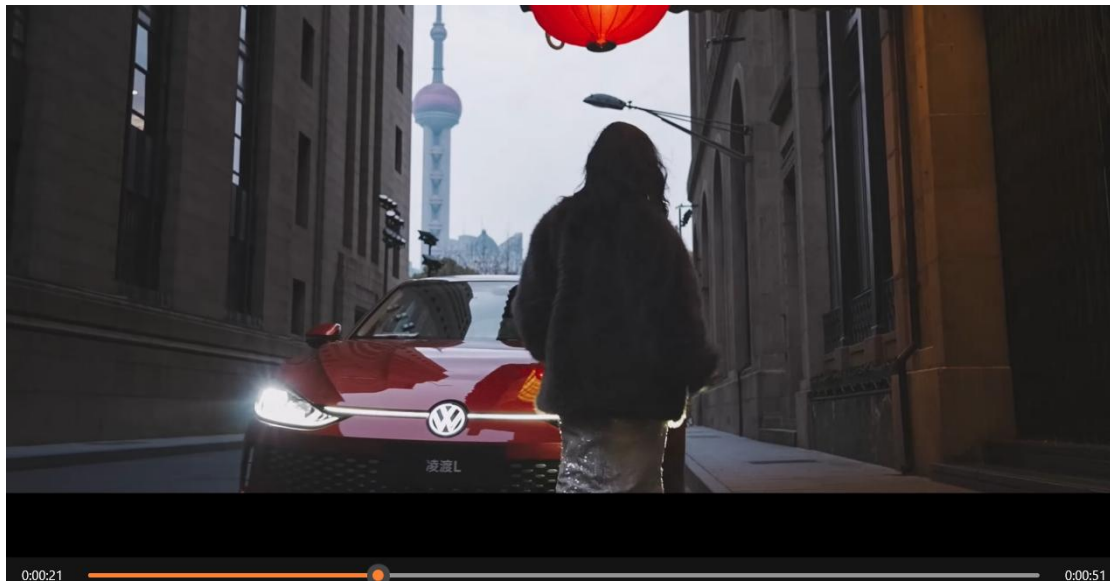


Fig. 34

In addition, color usage creates a sense of emotional continuity. Red, as the tone threading the entire video, appears repeatedly across urban scenes, which makes it a central symbol of emotional rhythms (see Fig. 35). It both signals the physical movements of people all across the city and actively engages with cultural spaces and emotional resonance. Such visual continuity leads viewers to savor the rich bonds that are passionate yet restrained, and personal yet shared between the creator and Shanghai.

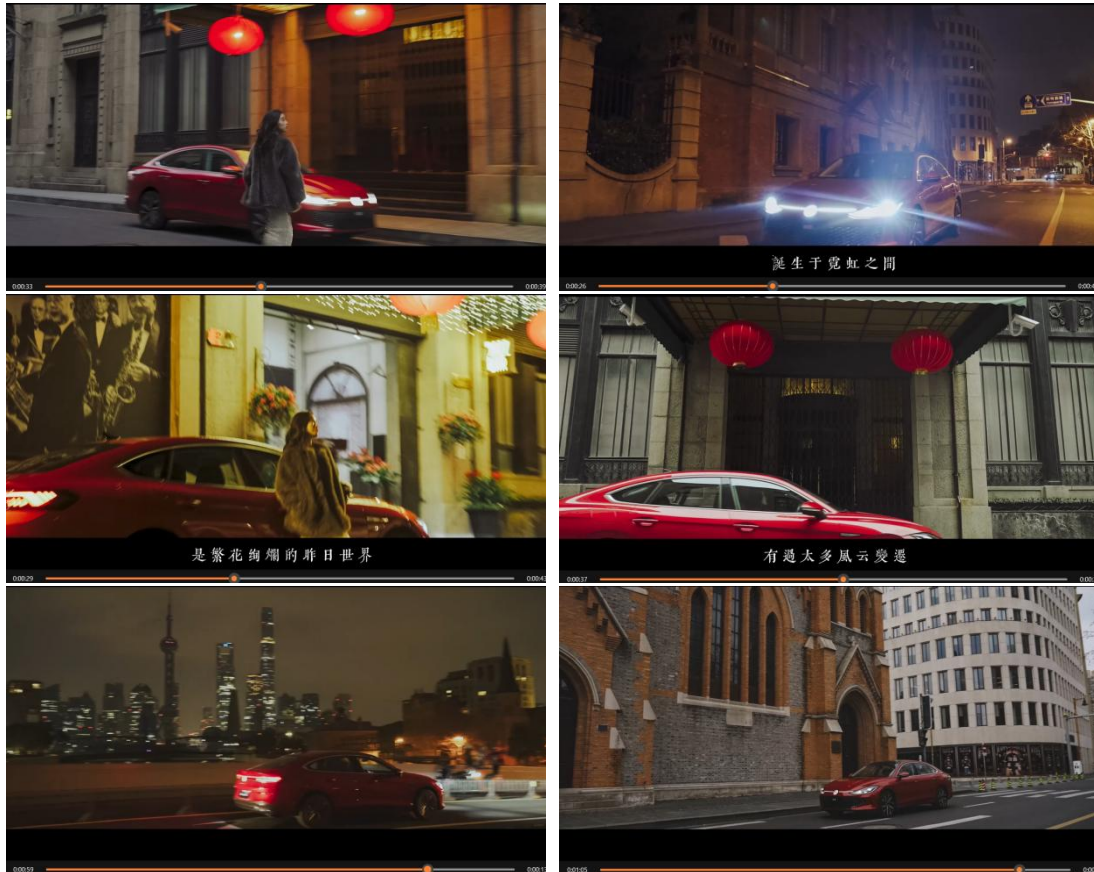


Fig. 35

Red concentrates the intensity of individual emotions, while dark tones highlight the city's historical depth. When combined, they construct an emotionally rich and warm image for the city. Shanghai is not just an object of observation but also a cultural space inviting emotional participation and reinterpretation under the creator's lens. Color's capacity to transcend visual information conveyance turns itself into a vector of intrinsic emotional expression, extending the city's image into a broader spiritual dimension.

The approach to color metaphor enables creators to infuse city images with cultural depth and spiritual meanings, where the portrayal of Beijing, Xi'an, and Shanghai shares common ground while retaining distinct identities.

During city image construction, creators use color as a visual presentation that emblemizes culture and memory, which goes beyond simple aesthetics and reflects deeper layers of urban character and cultural spirit. Visually impactful colors such as red, gold, and white are also invested with metaphorical meanings that imply authority, history, heritage, or vitality. These colors jointly help viewers intuitively grasp each city's cultural essence.

As depicted in short videos, color metaphors for Beijing focus on linking history with the present, applying a continuity of traditional colors to modern life to convey a sense of cultural stability and continuity. A muted, weighty historical atmosphere is cultivated by color contrasts for a balanced tension between tradition and modernity, hence expressing a powerful, restrained cultural identity of Xi'an. Color metaphors are used more emotionally and personally when depicting Shanghai's image. In all, creators weave colors into the infusion of city spaces and individual sentiments, constructing an inclusive urban image that blends memory, modernity, and personal experience.

Overall, color metaphors found in these short videos make contextual and symbolic expressions to raise the cultural richness of city images. Short videos thus emerge as a vital visual medium for embodying the city's spirit, unraveling a more layered and constructive insight into urban identity through color applications for the audience.

Conclusion

These short videos present colors from natural landscapes, cultural landmarks, and urban environments. Their creators have successively transformed spatial features, cultural meanings, and the city spirit into a persuasive visual language that helps viewers identify the city's exclusive traits and develop emotional connections with it when watching the short video. This context determines that color captures more than

just the city's visual reality. It also articulates symbolic and metaphorical meanings that enrich the cultural significance and aesthetic value for city image construction.

The color-powered expressions of Beijing, Xi'an, and Shanghai in short videos share visual rationales, such as the application of seasonal colors to intensify impressions of natural landscapes, or employing warm and cool tones to convey emotional ambiance and distinct, stylistic features. To illustrate, tones for Beijing emphasize solemnity and historical depth. The tonal approach for Xi'an lies in reflecting a blend of vitality and tradition with high-saturation colors. Softer tones are adopted in Shanghai's case to voice the warmth and multilayered nature of urban lifestyles. These discrepancies reveal each city's unique cultural foundation, with an emphasis on the diverse color narrative strategies in short video production.

Overall, color, as a highly condensed visual symbol in short videos, enhances the recognizability and communicative power of city imagery and emotionally resonates with viewers as well. Among the multiple pathways of city image construction, color includes emotion and memory in urban narratives and retains its important functions. The ensuing chapter will work to analyze the use of light and shadow in constructing the city images of Beijing, Xi'an, and Shanghai, respectively.

Chapter 4 The Role of Light and Shadow in Constructing City

Images

Introduction

This chapter examines how light and shadow are used as key visual components for constructing city images in short videos, and compares their differentiated applications in the cases of Beijing, Xi'an, and Shanghai. Light and shadow help convey emotions, shape the spatial features of a city, and bear cultural meanings. From natural light to various modes of lighting, the diverse usage of light and shadow enhances the visual expressiveness of the constructed image and infuses it with deeper layers of implications, including symbolic meanings.

This chapter begins by identifying the roles of natural light and shadow in shaping a city's overall visual character in short videos, which include both macro urban aesthetics and detailed portrayals of specific places. It then examines the influence exerted by atmospheric lighting on the multilayering of city places and mood cultivation, constructing further stand-out city images using techniques of brightness, color, and dynamic changes. Its subsequent section inquires into the symbolic implications of light and shadow, which lie in how they facilitate the communication of a city's history, culture, and emotions beyond undertaking the physical function. The combined effects of light and shadow nurture emblematic visual expressions in short videos.

Natural light and shadow in city image construction

In the new media form of short video, natural light and shadow are employed as crucial visual tools for shaping urban imagery. As noted by Zhang (2022), light and shadow, despite their origin in nature, function as a form of visual language. This

chapter thus further argues that creators are enabled by light and shadow elements to construct multifaceted city images. Short video creators can present the expansiveness of urban landscapes and reveal the intricacy of everyday details by capturing shifts in natural lighting, thereby making city images both magnificent and intimate.

In macro visual composition, changes in light and shadow intuitively empower the viewer to perceive and understand spatial depth, as well as scale and structure in the physical environment (Van & Keating, 2016). As shown in this study, contrasts between light and dark can enhance spatial layering through explicit distinctions between foreground, midground, and background, thereby reinforcing the frame's three-dimensionality. As demonstrated by Keating (2022), the fluidity of light and shadow infuses visual imagery with a dynamic quality that manifests how shifts in lighting influence viewers' temporal perception in short video production.

In addition, natural light and shadow are pivotal to shaping partial scenes. Compared to macro visual compositions, localized lighting is more detail-oriented, particularly underscoring the distinctive characteristics of urban places. Short video creators often apply techniques of contrasts, refractions, and shadow to enhance the texture of buildings, streets, and landscapes in order that viewers appreciate the refined beauty of a city through nuanced details.

Macro city aesthetics enabled by natural light and shadow

Short videos portray the instrumental roles of natural light and shadow in constructing macro city images. Natural light and shadow boost the spatial depth of urban environments and amplify distinct visual characteristics, developing a vast, stereoscopic, and multi-layered city imagery. The efficient capture of variations in light and shadow across varying weather conditions and periods of a day establishes a sense of order and structure in visual elements like architectural clusters, urban

landmarks, and street networks. This approach to visual depth elevation also reinforces the overall city image shown in short videos.

This section aims to dive into how natural light and shadow formulate macro urban aesthetic through short videos, covering their parts in defining spatial depth, enhancing visual effects, and strengthening the imagery coherence. Additionally, it adopts case studies to illustrate the contributions of natural light and shadow in constructing characteristic city images.

In the short video “I Have Never Been to Beijing”, the creator masterfully applies the interplay of natural light and shadow to depict Beijing’s city image as imposing, multidimensional, and richly layered. Apart from operating as natural phenomena, this interplay is found to be crucial to structuring the urban spatial order and augmenting visual depth, which imparts an impressive sense of proportion and aesthetic cohesion to the image. Dynamic usage of lighting outlines the spatial structure of Beijing and heightens the sense of depth and tension in the imagery.

The exquisitely distributed natural light and shadow within urban contexts is leveraged to highlight the contrast between Beijing’s natural landscapes and its built environment. The mountainous outskirts take on a dynamic visual appeal following light shifts from sunrise to sunset, during which the sunlight glows on the terrain to accentuate the stunning contour through interactions between brightness and shadow. Natural light and shadow shape a majestic natural landscape over the horizon by defining the outlines of mountains, as shown in the video (see Fig. 1). This effectively condenses the city’s geographical vastness into the short video format, leading the audience to immerse themselves in Beijing’s spatial composition.



Fig. 1

Beijing’s grand urban imagery is outlined by natural light and shadow as well. The luminance of sunrise over the skyline highlights the surrounding architectures, cultivating a more defined, layered silhouette in the city (see Fig. 2). Iconic high-rises filmed against the light emerge as dark silhouettes that stand in stark contrast to the luminous background. This technique generates a powerful visual impact and hence challenges conventional filming perspectives for the urban landscape, shaping dramatic imagery that enhances the video’s aesthetic appeal and shareability.



Fig. 2

Additionally, such light and shadow effects consolidate the diverse building surfaces in Beijing into dark outlines within the video frame. This approach abstracts the ornamental elements of building façades by further simplifying them into geometric forms. Elements such as glass curtain walls of high-rises, eaves of traditional buildings, and sharp edges of modern structures are all converted into distinct visual formats by natural light and shadow. As a result, the short video features the structured, orderly architectural style of Beijing. This visual treatment of simplifying

the frame composition guides viewers to more clearly grasp the urban layout and architectural features within the condensed viewing time.

This approach to natural light and shadow helps build up a macro visual aesthetic that seizes the city's vast spatial expansiveness, striking skyline, signature landmarks, and architectural geometry, thereby endowing Beijing's city image with a visually distinctive spatial charm.

In the short video "Xi'an, Please Take Another Look – A VLOG About This City", the skillful employment of natural light and shadow underscores the city's exclusive spatial layers and lends an organized, fluid beauty to its urban image. Besides stressing contrasts between the citywide architectural styles, natural light, and shadow heighten the perceived spatial depth, illustrating Xi'an's seamless blend of historical grandeur with modern vibrancy in the short video.

Wide-angle shots and side lighting are utilized to make natural light and shadow central to defining spatial relationships (see Fig. 3). Buildings that vary in forms and functions become distinctly visible under the sunlight, whose contours are clarified and contrasts emphasized with the interplay of natural light and shadow. This visual technique enhances the perception of Xi'an's layered urban fabric, setting architectural styles clearly apart from spatial elements. The video presents a more legible internal structure of the city using natural transitions of light and shadow, engaging viewers to experience Xi'an's complex urban composition and architectural diversity.



Fig. 3

The use of natural light and shadow also injects a stronger sense of depth into the city image on screen. Filming of the Xi'an City Walls presents full use of the contrast between light and dark under sun glows to deliver a layered visual effect (see Fig. 4). Sunlight dispersedly streaming through varied sections of the wall underlines the bricks' height variation and texture through shifts in light and shadow. The light-dark alternation also enhances the sense of continuity along ancient walls. This visual approach turns the historical structure of walls into part of the cityscape with higher profundity. Viewers can consequently sense the historical richness and spatial depth of Xi'an by watching vibrant movements of light and shadow in nature.



Fig. 4

Natural light and shadow are employed as instruments for constructing a distinct sense of layering in Xi'an's city image. Its spatial structure and historical identity are visually reinforced under wide-angle shots that showcase architectural relationships and light contrasts that highlight the continuity of city walls.

In the short video "SHANGHAI'S 24 Hours", the varied urban landscape is fully brought out thanks to the proficient capture of light and shadow for accentuating the city's spatial layering. This not only sketches the varying heights of architectural structures but also enhances Shanghai's signature visual identity as a skyscraper metropolis, rendering a highly immersive and magnificent image (see Fig. 5).

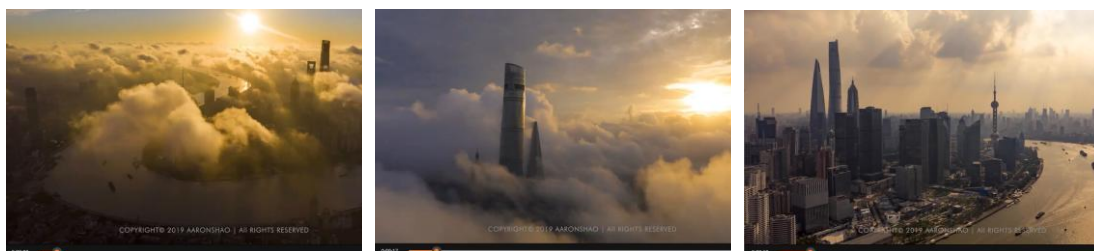


Fig. 5

The video kicks off with drifting morning clouds, shaping a mist-shrouded city ambiance. The cloud movements obscure the majority of the urban landscape, leaving the upper sections of iconic buildings the only visible scenery. Natural light and shadow play the key role of illuminating one side of exposed buildings as sunlight filters through clouds. The city's visual recognition is elevated by sharpening the architectural silhouettes with such a contrast. Despite substantial visual blockage of the cityscape by clouds, the precise use of natural lighting ensures instant recognizability of Shanghai even when outlines are just partially visible. This approach to city image construction highlights the spatial layers of the city through the light and shadow contrast. It also evokes the audience's visual memories of the city with landmark buildings, raising the immediate identifiability of the city image.

Natural light and shadow remain essential in shaping the city's aesthetic appeal. The video applies the Tyndall effect that appears when sunlight pierces through clouds and interweaves beams of light across the urban landscape. This introduces a rhythmic flow to the visuals, cultivating a simultaneously dreamlike and majestic image for Shanghai. The silhouettes of skyscrapers stand out impressively against light beams, while the shifting interplay of light and shadow between the clouds adds stunning depth and spatial complexity to the composition. Contrasts of light and shadow also assist creators with highlighting the city's stereoscopic structure, constructing the signature spatial allure of Shanghai in short video creation.

In this case, the creator presents Shanghai's visual layers at various times throughout the day. Natural light and shadow increase architectural recognizability by imparting depth to spatial relationships during city image construction. As a result, the short video conveys the key city features within the restricted video time. Regardless of depicting towering buildings emerging above the clouds, or the vast urban landscape bathed in light transitions, the representation of natural light and shadow presents Shanghai as a modern metropolis imbued with a refined spatial aesthetics.

The three city case studies boast a number of shared features in constructing macro city images with natural light and shadow in short videos. Firstly, natural light and shadow, as primary tools for location denotation, are widely used to enhance the sense of layering and structure in city visuals, making buildings, landmarks, or natural scenery more visually striking. Secondly, they optimize the sense of order and aesthetic quality of the frame. Techniques such as light-dark contrast and light movement are used for constructing spatial logic and visual rhythm. This all instantly engages the audience to acquire clear-cut insights into the city's spatial layout.

Despite similarities in strategies, the visual presentation of each city significantly varies. Natural light and shadow are applied to emphasize structural grandeur and build an aura of solemnity in the video of Beijing. The video constructing Xi'an's image lays stress on the historical depth of places, blending ancient buildings with the natural environment to present a visual structure fusing history and modernity. In contrast, Shanghai's video case prioritizes the interaction between towering buildings and shifting lights.

Natural elements like drifting clouds and sunrise help construct its modern, dynamic aesthetic. It is noted that short video creators adapt their use of natural lighting to accordance with specific contexts in each city, thus cultivating in the city image a more intense localized expression and recognizability. In a word, this reflects discrepant spatial forms and cultural identity among the three cities.

Refined aesthetics through natural light and shadow

Apart from fashioning the overall structure of the city image, natural light and shadow also help reveal the delicate details of micro scenes, investing the city imagery with visual richness and aesthetic depth. The meticulous lighting control by creators

enlivens the diversity and aesthetic particulars of local places, helping to build a city image brimming with a strong sense of depth and layering. This observation aligns with Zhan's (2024) discussion on the visual functions of natural light and shadow.

This section probes into natural light and shadow to shape localized places and construct a cultivated urban aesthetic in short video production. It examines how light and shadow enhance the expression of spatial details and contribute to a city's distinctive visual appeal with their variations. The diversified different manifestations of natural light and shadow are further covered with city case studies.

The short video "The Most Characteristic Beijing" is characterized by the precise control of natural light and shadow to vividly showcase the exquisite aesthetics in Beijing hutongs that feature spatial depth, a sense of order, and a lived-in atmosphere. Natural light and shadow highlight intricate details to make objects, scenes, and people within hutongs more visually distinct. Dual emphasis on static and dynamic elements enabled by light and shadow communicates a wealth of visual information within a limited viewing period, effectively capturing the characteristic texture of hutong spaces.

In the opening scenes, sunlight tickles through leaves and casts dappled shadows on the ground (see Fig. 6). This interplay amplifies spatial depth in hutongs and forges a distinct contrast between light and dark. The left side of the frame records the well-defined contours of a parked bicycle, neatly stacked bricks beneath a window, and a cyclist traversing the street center as natural light and shadow mingle, thus featuring more stand-out spatial relationships. The limited natural light in hutongs results in structured and harmonious illumination, establishing a sense of order within the frame. Natural light and shadow, by reinforcing the structural composition of hutongs, draw viewers to perceive daily hutong lifestyles through the meticulous spatial arrangement.



Fig. 6

The lighting strategy delivers a similar effect in another set of shots (see Fig. 7). As sunlight casts glows on a bicycle parked against the wall, the contour of its metal frame appears more distinct under the light-shadow contrast in the frame, thereby enhancing the spatial dimensionality. Similarly, the light highlights the texture of agricultural products hanging outside a window for air and drying, demonstrating how light and shadow portray the physical properties of objects.

The scene where an elderly person is basking in the sun at a street corner tenderly shapes an interplay between the figure and surroundings to build a more invigorating image of the figure. Details of objects, as underscored by light and shadow, represent the creator's intention to draw the audience's attention to easy-to-overlook daily components, thus producing a more delicately and expressively filmed video.



Fig. 7

In the short video “The Only City Worthy of the Glorious Tang—Chang’an”, natural light and shadow are adeptly applied by the creator to characterize the elegant aesthetics of Xi’an’s streets and alleys, presenting distinctly layered street scenes, crowds, and historical architecture with strong visual clarity. Upon shaping the cityscape structure, they also strengthen the expression of finer details for the representation of Xi’an as a lively city embodying profound historical texture. The precise regulation of natural light and shadow enhances the visual tension in the urban landscape, effectively conveying an abundance of city details in a narrow time window.

When watching the street scenes, the audience can visually follow the creators’ movements through the bustling downtown area where the people weave through the streets (see Fig. 8). The faces of passers-by are illuminated by the glowing sunlight, which further outlines their facial expressions and renders distinctly recognizable figures using the contrast of light and shadow. This strategy illustrates a more vivid presence of people and further realistic, layered street scenes. Additionally, as seen from shots of storefronts along the streets, sunlight accentuates the vibrancy of shop signs, intensifying the visual impact and enriching the representational depth of the city’s commercial places.

The variations in natural light and shadow cultivate details with greater expressiveness, leading to viewers’ intuitive perception of the spatial layout and commercial ambience of Xi’an’s streets by the audience when appreciating the video.



Fig. 8

In another set of shots, a stable visual structure between Xi'an's historical architecture and human figures is presented with the application of natural light and shadow (see Fig. 9). The strategic use of lighting angles ensures shots featuring the Giant Wild Goose Pagoda operate as follows: figures are set in the foreground, pagoda appears in the background, and Xuanzang's statue are explicitly placed within the frame. Such usage of light and shadow puts added emphasis on subjects portrayed in the frame. The short video thus creates a more straightforward context where viewers grasp the city details that the creator has observed.



Fig. 9

In the short video “Shanghai Alleyway Nostalgia”, natural light and shadow are used to display the elaborate details in Shanghai’s alleyways, shining a light on everyday life by introducing a distinctive sense of depth. They both illuminate objects within the place and establish an orderly visual cadence through the interplay of brightness and darkness, enriching the expressiveness of daily life. This results in a diversified, in-depth representation of daily scenes. The alleyway, being a narrow but multi-layered urban place, is applied as a visually characteristic and rhythmic composition that enriches the spatial structure of the short video.

As sunlight pours down on the drying laundry, shifting light accentuates the clothing’s texture (see Fig. 10). A line of laundry hanging on the clothesline is a common scene in life. However, the framing of natural light and shadow can turn them into visual symbols that bring a stronger sense of order to the spatial structure. Applying light and shadow captures vibrant visual relationships within everyday scenarios, bringing subtle movements to spaces that would otherwise appear static.



Fig. 10

Clotheslines extending from residents' windows cast intersecting shadows onto the walls and architectural structures in the sun (see Fig. 11). This layering of light and shadow clarifies the spatial relationships within the alleyway, while the contrast between clotheslines and buildings accentuates the compositional depth. The sunlight, when seeping through narrow passageways, falls on windowsills, door frames, and walls, introducing a rhythmic interplay of light and shadow. This dynamic variation in brightness structures the alleyway interiors with a sense of depth and dimensionality, thus designing a more visually engaging spatial composition.



Fig. 11

In another shot sequence, soft lighting drifts through a chair deep within the alleyway. An elderly resident sits on the chair, attentively reading a newspaper (see Fig. 12). The transitions in light and shadow lend a serenity to this everyday moment. As gentle shadows cast across the newspaper, the page-turning process becomes more pronounced under the interplay of natural light and shadow that outlines the figure. This imbues the everyday scene with a sense of stability and visual order in its composition. This visual expression features natural light and shadow as key elements to frame structuring, adding delicate layers of detail into routines in the alleyway.



Fig. 12

The three cities present similarities in how they express the refined aesthetics of natural light and shadow in short videos. Creators often combine natural lighting with detail-oriented framing to fill micro-scenes, everyday objects, and human figures with a stronger sense of depth and texture on screen.

Ranging from daily scenes in Beijing hutongs to the people movements along Xi'an's busy streets, or the common life in narrow alleyways of Shanghai, proficient use of natural light and shadow is found everywhere. This highlights details and vibrant, distinct appearances of objects and individuals, and even organizes the visual composition in these smaller city spots. They make the visual structure orderly and infuse the sense of narrative potential into the visual layout. These shared strategies of natural light and shadow reflect creators' endeavors to help viewers directly relate to the impressive textures of city life.

Meanwhile, each city differs in the application of natural light and shadow as per city-specific features. The short video for Beijing emphasizes the structured, order-centered cadence of urban life. Xi'an's video case constructs layered visual

relationships that guide viewers to discover the city in depth. The short video portraying Shanghai takes an approach of reinforcing everyday aesthetics and creating a serene, warm visual atmosphere.

Atmospheric light and shadow in city image construction

Light and shadow are also key to building the urban atmosphere. Atmospheric lighting concerns not only light elements but also the distribution of light and shadow, their movements through city places, and interactions with surroundings, all of which facilitate the production of visuals that are more expressive in depth and mood.

The arrangement of atmospheric lighting performs a key function in organizing spatial structures within short videos. Adjusting angles and distribution of light and shadow creates clearer layering across areas, and more and easier-to-achieve visual emphases. (Wang, 2022).

Transitions between brightness and darkness affect the visual relationship connecting buildings with streets. In short video production, the height, position, and coverage of light sources collectively shape the degree of openness, creating distinct contrasts between nearer and farther areas. According to Wang (2024), when light and shadow intersect, objects are endowed with a more pronounced visual depth. Short video creators can also use light sources of varying intensities and colors to reinforce the structural sense of city places, presenting a layered spatial layout within a two-dimensional image frame.

Changes in atmospheric lighting affect the intuitive experience of the city's ever-changing moods. The tonality of lighting, the light-dark contrast, and rhythmic variations, as different applications of light and shadow, all shape the overall atmosphere, displaying a city image of varying visual qualities. In smaller urban

spaces, light and shadow help evoke specific moods to cultivate more recognizable on-screen city images. Adjustment of lighting in short videos enhances the entire visual impact for a more expressive, signature city image construction.

Additionally, atmospheric lighting shapes the emotional tonality of a city. Brightness, color temperature, and variations in light and shadow influence the overall ambiance, arousing varied visual perceptions in the city's image. Softer lighting tends to convey warmth and tranquility, while cooler tones or high-contrast lighting can reflect the city's dynamic rhythm or historical depth. Short video creators alter the layering of light and shadow as they are not content with merely depicting urban elements, and apply atmospheric lighting to enhance the emotional engagement and deliver a more engrossing visual experience for viewers.

This section examines the application of atmospheric lighting in short videos, focusing on its role in shaping spatial depth and emotional ambiance. It also penetrates into constructing more expressive, compelling representations of city imagery through light and shadow.

Spatial layering through atmospheric light and shadow

Atmospheric lighting functions as both a means of illumination and a key visual element for constructing the layered structure of a cityscape. Unlike natural light and shadow, atmospheric lighting primarily originates from artificial sources, including neon signs, decorative lights, and street lamps. It produces light-dark contrasts that further clarify spatial layers with variations in colors, brightness, and projection angles (Zhao, 2021). It also emphasizes the boundaries of places for registering a higher sense of spatial hierarchy.

Bright areas are naturally eye-catching, while gradual transitions between light and shadow forge smoother spatial flows, enhancing the depth of city images. As a result, creators bring out more visible spatial layers and deepen viewers' comprehension of the city structure by arranging and combining these light sources where appropriate.

This section is dedicated to how atmospheric lighting constructs spatial depth in short videos, encompassing analyses of its influence on spatial organization and how variations in lighting make urban imagery more expressive.

In the short video “Finally Understand Why Everyone Has an Obsession with Beijing”, atmospheric lighting is utilized to highlight the functional divisions within Beijing's urban landscape, enabling visually-distinct spatial hierarchy through contrasts of light and shadow (see Fig. 13). A single fixed shot captures the natural transitions from dusk to night as a temporal axis, where the gradual shifts in lighting and atmospheric illumination reveal functional characteristics of different city zones. Light and shadow divide the urban place in the visual composition, in which scattered, gently glowing residential areas contrast with the concentrated, striking illumination of urban landmarks, explicitly distinguishing the hierarchical relationships among different functional zones.



Fig. 13

The video begins with a dusk view of the city. Following the fading of natural light, atmospheric lighting gradually takes over as the primary source of illumination. The lit-up Temple of Heaven and Tian'anmen Square grow particularly prominent against darkened surroundings. Such lighting effects, which serve as spatial markers throughout the video, shape significant visual contrasts to reinforce the spatial identity of these iconic historical and cultural sites and turn the sites into visual focal points. They are central to the layering of the cityscape as well. The light-dark contrast adds visual depth to the buildings whose roles in the city's overall structure are further characterized.

The background features how scattered points of light illuminate the residential area in the distance, forming a stark contrast with the bright, concentrated lighting in the landmark districts. The softer and further dispersed lighting in residential areas supports these areas with a gentler visual imagery. In contrast, signature areas become more visually striking against the night sky due to their centralized, intense illumination. This contrast in light and shadow enhances the cityscape layering, thus defining clearer spatial boundaries between functional zones.

The single-frame composition smoothly captures the fluidity of city lighting. As night deepens, atmospheric lighting grows into the visual element dominating and outlining the spatial hierarchy across the city. This minimalist approach conveys rich information. The interplay of light and shadow stresses the prominence of core city areas, enabling the audience to intuitively grasp the functional divisions in Beijing by identifying variations in brightness and darkness. This representational mode raises the three-dimensionality of the city image construction while showcasing the richness of spatial layering within the video's short duration.

In the short video “Tonight, the Xi’an City Walls Spring Festival Lantern Show Officially Lights Up”, atmospheric lighting enriches the city’s image by presenting a distinctive sense of spatial depth through festive illumination. Centered on the Spring Festival lantern show, the video forges an intimate spatial bond between Xi’an City Walls and festive lanterns by orchestrating light and shadow, which provides the city with a stronger celebratory mood. Contrasts between light and dark, and variations in depth of lighting, all introduce a richer sense of spatial layering to the video.

Dragon-shaped lanterns are arrayed along city walls, standing in parallel to ancient architecture (see Fig. 14). The application of atmospheric lighting accentuates the walls’ silhouettes against the nighttime sky, and highlights the solidity and texture of the stonework. How light and shadow emphasize the three-dimensionality of the architecture makes the spatial layers of the city walls more distinct at night. Viewers can sense the grandeur of the Xi’an City Walls through shifting lighting effects. In visual expression, brightness, color, and arrangement of decorative lights, combined with architectural features, reveal an exclusive cultural appeal shaped through light and shadow in the short-video-based city image construction.



Fig. 14

The continuous stretch of lanterns into the night imparts openness and a sense of depth to the city's spatial representation. The lantern arrangement reinforces the linear extension of city walls, adding visual depth within the limited time frame, and enhancing the three-dimensionality in city spaces. The contrast between illuminated lanterns and the dark background more clearly outlines the city's spatial structure, therefore reinforcing the audience's visual perception.

The light and shadow orchestration in the short video seamlessly integrates the atmospheric lighting of Xi'an City Walls with festive glows of Spring Festival lanterns, showcasing the diversity of spatial layering within the cityscape. Blending decorative lights with city walls highlights the local culture around the city. The contrast of light and shadow clarifies the relationships between architecture, lighting, and space. The alternations between brightness and darkness nurture a layered visual experience. In a word, festive, atmospheric lighting helps Xi'an take on a distinctive city image.

In the short video "Shanghai Impressions: Aerial View of the Glittering Bund at Night", atmospheric lighting spotlights the vibrant, dynamic image of Shanghai. Filmed from an aerial perspective that glides over the city's high-rises, the video captures the dense clusters of commercial buildings and structured citywide spatial design. Light and shadow enhance the clarity of its structural composition, making Shanghai's towering skylines stand out against the illuminated cityscape.

Decorative light strips on skyscrapers and interior glows from glass curtain walls collectively define the city's vertical spatial attributes (see Fig. 15). The distribution of light and shadow adds pronounced spatial layers to the night scene, elevating the stereoscopic nature of the architectural styles. Variations in lighting tones and brightness across buildings further trace the depth of the city's spatial connections, reinforcing the structured composition of its commercial districts within the limited video duration. Movements of aerial shots enhance the interplay of light and shadow, outlining Shanghai's striking silhouette with greater clarity.

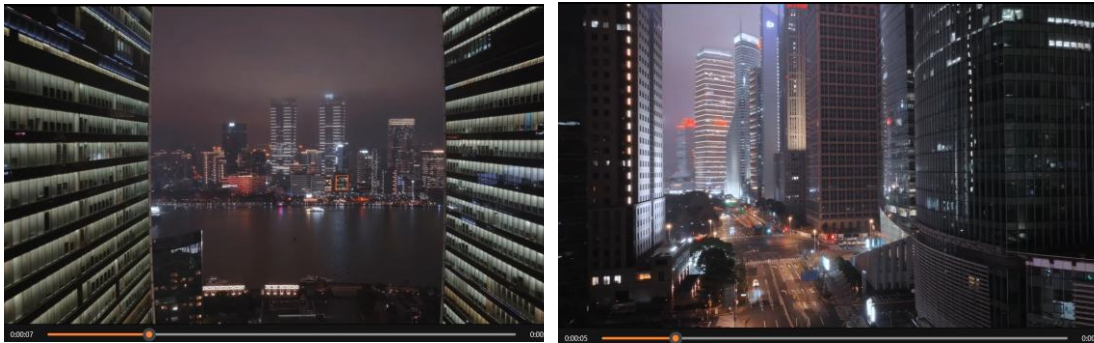


Fig. 15

The aerial perspective applies vehicle light as another key element for light and shadow in the imagery as well (see Fig. 16). The continuous streams of headlights and taillights create dynamic light trails, contrasting with the static illumination of skyscrapers to make a multi-layered spatial composition. The lighting on the roads, as a kind of light and shadow distribution, sketches the city's transportation networks and further reinforces the spatial organization. The light-dark contrast better clarifies spatial relationships, with a depiction of Shanghai's densely packed skyscrapers and well-ordered streets within the nighttime cityscape in the video.

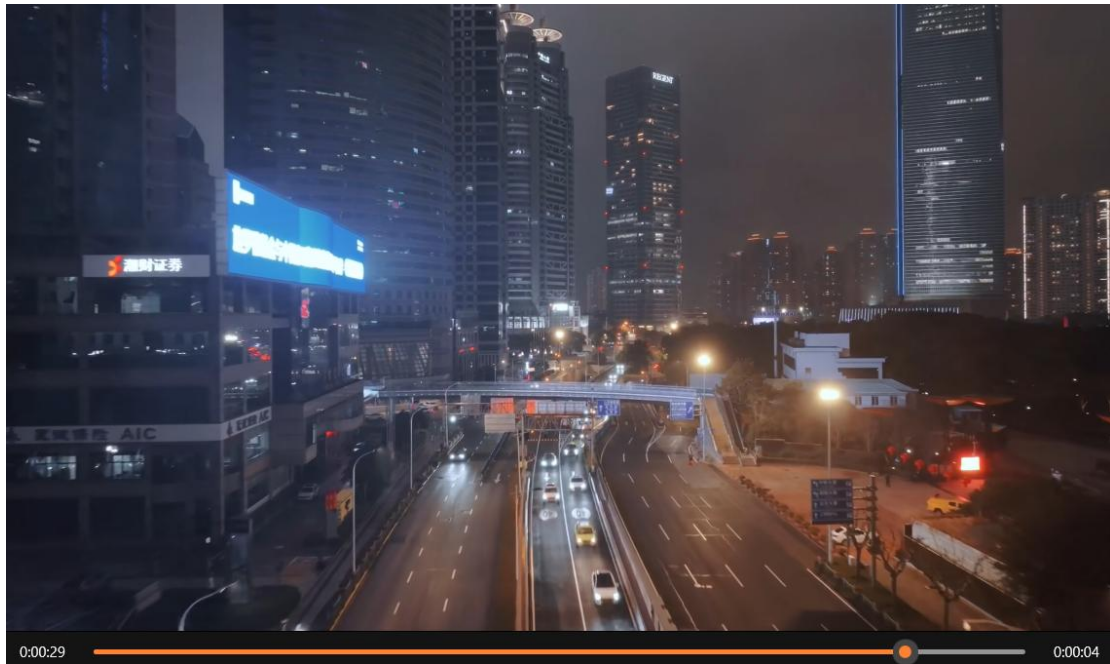


Fig. 16

By combining aerial perspectives with atmospheric lighting, the short video presents commercial districts in Shanghai with a rich spatial depth. The lighting configuration for buildings outlines the city's silhouette, while the dynamic orchestration of vehicle lights further enhances the perceptual depth, developing a more defined city image when applying light and shadow. Lighting, apart from accentuating the spatial characteristics of the city, constructs an organized, lucid city image through the layered distributions of varying light sources.

In the three short video cases, atmospheric lighting, as a controlled visual source, is widely used to recreate the visibility and spatial depth of the urban nightscape. Compared to the unpredictability of natural light, atmospheric lighting is more controllable and flexible in design. Creators draw viewers' visual attention by arranging light sources, adjusting brightness, selecting colors, and intentionally framing the illuminated city spaces against the dark background (Malkiewicz, 2012) to shape an easy-to-identify visual structure.

Vivid examples include: light-dark contrast employed to show functional zoning in Beijing's case; a blend of festive lights and historic architecture in the video for Xi'an; or the manifestation of spatial order through commercial lighting transitions in Shanghai's video. They all elucidate the integral role of atmospheric lighting in shaping the visual presentation of city nightscapes.

While all three city-based videos construct spatial layers with atmospheric lighting, their visual strategies and points of emphasis differ.

In terms of Beijing's video, it works to separate functional places in visual terms. The creator prioritizes shaping a structural layout for functional zones via atmospheric lighting. In Xi'an's video case, colorful lights for celebration turn historic architectures into emotional symbols that bring the local culture allure to the nighttime cityscape. Shanghai's video example highlights the city's dynamic structure. Light strips surrounding skyscrapers and the fluid trails of traffic collectively build a vertically flowing space, creating a fast-paced, high-density urban impression. As seen from these differences, atmospheric lighting both defines the contours of spatial structures and manifests creators' employment of lighting strategies in alignment with each city's cultural and sensory character.

Atmospheric lighting and expression of city ambience

Atmospheric lighting is fundamental to shaping the mood of a city, whose brightness, color, projection method, and distribution of light and shadow all influence perceptions of the overall ambience (Watkins, 2022). These attributes visually add emotional characteristics to the city image. Strategic application of atmospheric lighting, as enabled by the vibrancy of light and shadow interplay, presents cities with distinct emotional tones, thereby drawing viewers to experience the unique ambience (Huang, 2021) through the visual production of short videos.

This section investigates the detailed construction of urban atmosphere by atmospheric lighting through short video cases, such as by examining the roles of brightness, color, and light and shadow distribution in shaping the visual atmosphere. Its case studies delve into the ways of enhancing the expressiveness of city imagery through light and shadow treatments.

In the short video “Click to Experience the Full Festive Atmosphere of Fireworks at Qianmen”, the creator displays a warm, lively nighttime ambience of Beijing hutongs (see Fig. 17) through atmospheric lighting. The interplay of light and shadow, in addition to shaping the spatial depth, arises as a compelling instrument of emotion communication that makes the audience feel at home. The video, by regulating variations in brightness, projections enabled by different materials, and contrasts in lighting, constructs a richly layered nighttime cityscape, engrossing viewers in admiring the unique atmosphere of nights in Beijing hutongs.



Fig. 17

The video unfolds with a shot in which an elderly citizen is busy cooking in the kitchen, whose actions are glimpsed through a window. The kitchen light serves as the only light source in this night frame, casting the elderly figure’s silhouette against the glowing light. The light-dark contrast creates layered divisions within this particular location. Soft lighting enhances the visual perception of a warm, cozy indoor space. This composition doesn’t just employ lighting for illumination but also

uses its contrast with the surrounding darkness to render a striking visual impact, further enlivening the familial nighttime in hutongs.

As the camera moves into the kitchen, the lighting atmosphere changes. Bright light illuminates the food, stressing its rich colors and enticing nature. Meanwhile, the steam rising amidst chilly night air is especially prominent. Gentle lighting, whose visual texture is inviting and warm, diffuses through the steam. This light and shadow interplay helps with cultivating a more authentic vibe of familial gathering as well as reinforcing the everyday glamour of hutongs.

The ending shot shifts toward the semi-transparent window curtain, capturing a view of a family get-together inside the room. The translucent visual fabric diffuses the interior lighting to create a soft glow, while the silhouettes of family members overlap at the intersection of light and shadow, elevating the compositional depth. Light and shadow again undertake functions beyond lighting sources, working as spatial separation markers by creating a contrast between indoor and outdoor areas. This lighting treatment creates cozier and more emotionally engaging family gathering vibes.

The short video relies upon atmospheric lighting to visually recreate the nighttime family ambiance in hutongs. Light and shadow are instrumental in shaping the emotional tone of the imagery. Through the contrast of dark silhouettes outside the window, the bright kitchen lighting, and the soft glow diffused through the curtain, the video portrays a touching nighttime in Beijing. This lighting-inspired, splendid imagery again engages viewers to experience the warmth and emotional depth of hutong life.

In the short video “Changxiangsi, in Chang’an”, localized atmospheric lighting crafts a distinctive cultural ambiance, transporting viewers back to poetic, picturesque days

in the past of Chang'an (former name of Xi'an). Beyond illuminating the scenes, atmospheric lighting enhances the city's cultural depth by deploying various strategies and layered variations in lighting. The blended history and modernity within the imagery construct a city image rich in cultural symbolism (see Fig. 18).



Fig. 18

As observed from the visuals, atmospheric lighting is substantial to evoking a historical context. Under the interplay of light and shadow, actors and actresses portray historical figures such as Xuanzang and Li Bai, whose emergence gets viewers to sense the city's cultural essence through the projection and transformation of silhouettes. The soft light transitions help highlight figures' contours and the immersive quality of the short video scenes. This adds to the cultural expressiveness of the visuals. In addition to shaping the visual appearance, light and shadow are key to constructing the holistic atmosphere in the video, in which viewers recognize the unique cultural texture of Xi'an.

This cultural representation is further augmented by artful lighting design. The lanterns crafted in traditional methods provide major lighting sources, whose large quantity imparts ample brightness to the scene and thus visually enhances the historical elements. Such pragmatic lighting treatment also stands as an integral cultural symbol within the video, taking viewers to acknowledge the city's cultural depth when appreciating the fluid movements of light and shadow. The layered colors and brightness contrasts facilitate a more immersive and culturally relatable representation of the city's image and cultivation of the visual depth.

Atmospheric lighting introduces historical profundity to the scene by enriching the already immersive experience in the video, and expressing the fascinating urban culture with the adept usage of light and shadow. The dynamic interplay of lighting strategies helps construct a rich cultural imagery and presents Chang'an's poetic, narrative-focused visualized image. Viewers are dynamically drawn into the historical appeal of this ancient capital and experience its cultural legacy in a visually compelling way during alternations between light and shadow.

In the short video “The Night View Bus Tour of the Bund—My Ultimate Guide!”, the creator enlivens Shanghai's modern city image featuring a vigorous nighttime cityscape by applying atmospheric lighting. Filmed from the perspective of a sightseeing bus, the video captures the dense illumination as the dominant visual tonality along both sides of the landmark Huangpu River. Atmospheric lighting, aided by changes in color and brightness, delivers a strong visual impact at the intersection of light and shadow for building the nightscape image.

The sightseeing bus traveling across the city occupies the core imagery of the short video as the visual thread for viewers, is accompanied by its surrounding atmospheric lighting that continues to shift along the designated route (see Fig. 19). The radiance of brightened buildings, reflections from glass curtain walls, neon streetlights, and waterfront landscape lighting together create a dazzling, fluid visual spectacle. The layering of colors and rhythmic changes in lighting further enhance the immersive identity of the urban nightscape, shaping a more vivid, tangible city atmosphere.

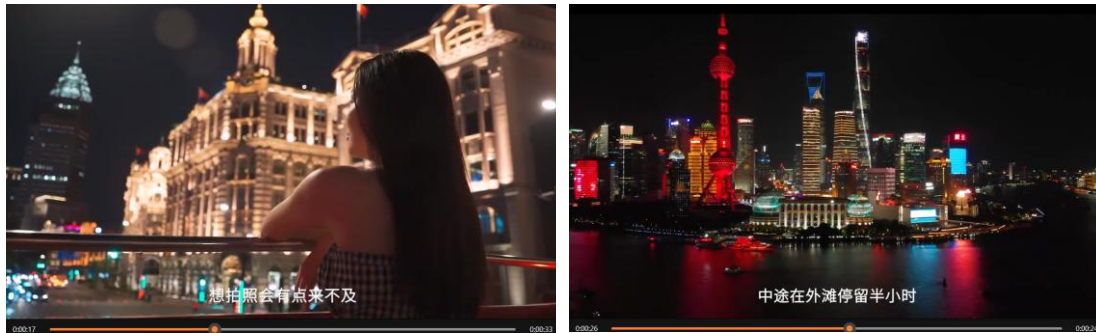


Fig. 19

Large electronic billboards and illuminated advertisements over skyscrapers play pivotal roles in shaping atmospheric lighting (see Fig. 20). As the sightseeing bus moves forward, color variations create an ever-shifting nighttime ambience abundant in dynamism. These bright light sources shape a more visually imposing nightscape and further animate commercial districts, making light and shadow essential to expressing the vibrancy of urban commerce. The rhythmic transitions of lighting, in sync with the dynamically updating content of the billboards, further highlight the city's modern texture, allowing viewers to live through bustling engagements in commercial areas of Shanghai through the video.



Fig. 20

Through the crafting of atmospheric lighting, the short video effectively constructs a vibrant and urban nightscape of contemporary Shanghai. The multifaceted interplay of skyscraper lights, commercial billboards, and neon streetlights infuses the urban imagery with constant streams of energy, strengthening Shanghai's visual identity as a global metropolis.

The short video cases from the three cities all spotlight the significance of atmospheric lighting in cultivating the holistic urban vibe. This lighting approach prioritizes more on engaging viewers' emotional perception and sensory immersion. Through appropriate alterations in the color temperature, brightness, and distribution of light and shadow, these city visuals are produced to convey emotional warmth, cultural depth, or urban vibrancy.

Light and shadow, beyond functioning as illumination instruments, are central to shaping the urban ambiance and guiding viewers to perceive the city's imagery through visual experience. Mellow lights that stream through the windows of Beijing hutongs at night, lighting details like cultural silhouettes lit by traditional lanterns in Xi'an, or the shifting neon lights visible through a car window in Shanghai, all transform physical places into emotionally relatable scenes.

In short videos, each city's image construction varies in the employment of atmospheric lighting. The lighting highlights the warm, intimate nature of everyday scenes in Beijing's video case. Warm tones and localized lighting help strengthen the familial and community sense in nighttime scenes. Xi'an's video applies historically motivated lighting to create a temporal sense of cultural continuity, setting the city as a place rich in narrative and tradition. In contrast, lighting in Shanghai's case emphasizes the cadence and energy of a modern city. Neon signs, illuminated buildings, and flowing colors jointly create a lively nightscape. These differences

reflect the varied styles of shaping each city's atmosphere and using lighting design to recreate that atmosphere into a visible emotional experience.

Symbolic meanings of light and shadow in city image construction

Light and shadow, as significant visual elements for city image construction, embody emblematic meanings that lend cultural profundity and emotional resonance to the imagery. Symbolic visual effects are expressed through lighting techniques such as illumination styles, contrasts between light and dark, and the use of color (Ford, 2020), transcending physical locations and thus symbolizing specific cultural values, historical narratives, or social identity in city image construction. For example, variations in light and shadow can strengthen symbolic connotations, embodying deeper hierarchies of cultural significance within the image by portraying buildings, streets, and natural landscapes.

Different lighting atmospheres in the video, empowered by the key medium of light and shadow in urban environments, can elicit a range of emotional responses from viewers (Wang, 2013). Movements of light and shadow dually affect the perception of exact places in the cityscape and establish the emotional tonality for the narrative, constructing a more engaging, expressive city image.

The following discussions will inquire into the symbolic visual effects of light and shadow and their roles in emotional expression, as well as the application of lighting in short videos to endow cities with deeper visual meanings and emotional value. This will make cities both physical landscapes and, more importantly, part of the emotional experience and cultural bonding.

Symbolic visual effects of urban light and shadow

According to visual semiotics, besides serving as physical elements, light and shadow also embody emblematic meanings. The contrast between brightness and shadow can highlight the spatial structure and intensify the visual impact of the city. Soft or focused lighting can draw the viewer's focus to specific visual points, investing the image with a distinctive sense of meaning (Wisessing et al., 2020). As depicted in this study, various lighting strategies help shape a city's visual style, empowering it with stand-out aesthetic features and strengthening the viewer's impression of the city's image.

The following discussion will focus on how symbolic visual effects of light and shadow influence the city image construction, and exemplify how short video creators use lighting to construct city images with symbolic significance through specific city examples.

In the short video “The Red Walls and Green Tiles of the Forbidden City, Holding the Years of Countless Women”, light and shadow are applied to forge a symbolic visual effect, visually representing the passage of time and spatial confinement. The movement of light and shadow, combined with thoughtful composition, can enhance the visual depth of the image and infuse it with deeper symbolic connotations, hence presenting the city's characteristic visual identity (see Fig. 21).

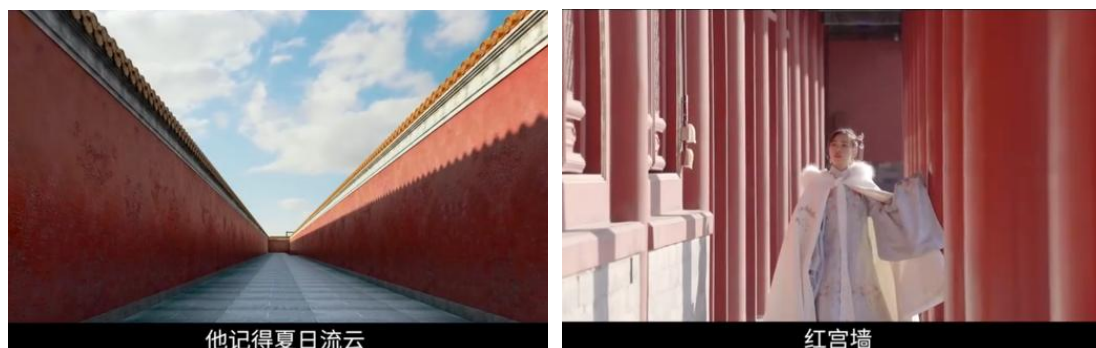


Fig. 21

The use of dynamic light and shadow strengthens the viewer's temporal perception. The video uses slow-motion footage to capture the movements of shadows across walls of ancient palaces. While the objects in the architecture remain static in the frame, shadows slowly shift across the surface, cultivating a sense of time in motion. This shift in light and shadow symbolizes the passage of time, moving beyond a static spatial presentation and helping viewers sense the time-shaped experience of places.

The composition of light and shadow further builds symbolic visual connotations. The frame, situated in a corridor of the Forbidden City, features a figure standing in the center. On the left side stand wooden pillars, whose shadows are cast in neatly vertical bands onto the wall on the right under the sunlight. The scene as a whole forges an enclosed visual space. This image structure, shaped by the play of light and shadow, adds tension and a sense of confinement to the visual composition through which the creator manifests a particular historical narrative in the city image by introducing symbolic expressions.

The synergy of light and shadow transitions and thoughtful framing presents a visual image of both time and confinement within Beijing's historical sites in the short video. Light and shadow arise as the key visual elements bearing historical atmosphere and symbolic meanings, designing a more layered and meaningful visual language for the city.

In the short video "Xi'an City Walls", the distinctive light and shadow of a rainy day symbolically construct visual effects to shape a blend of history and modernity in the image. The damp environment, when applied with light and shadow, features distinct layers through scattered lighting, reflections, and the movement of raindrops, all adding depth to the symbolic meanings. These elements visually express the perception of advancing with the times while holding onto its historical roots within the city space.

The dispersed lighting on a rainy day softens the sharpness of unobstructed illumination, imbuing a hazy and misty quality (see Fig. 22) to the entire imagery. Gentle lighting projected on city walls embodies a heavier and more solid visual identity. The wet surface reflects a faint glow, adding richer texture to the architectural details. This approach to lighting blurs the temporal sense so that the visuals carry both historical profundity and modern character. The soft transitions resulting from scattered light bring about a low-contrast visual style, building the tranquil yet intense atmosphere as the emotional thread throughout the short video.

The reflection of the creator as the figure running in this video is central to reinforcing symbolic meanings. Dressed in traditional-style clothing, the creator's movement of running along damp city walls has been captured using the low-angle tracking shot that interweaves her real imagery and reflection. The reflection's brightness and shape change following each step forward. This treatment of light and shadow adds symbolic depth by turning the person-reflection interplay into a visual metaphor that voices the historical and modern overlap. The blurred, distorted reflection further enhances the visual tension, delivering multi-layered implications despite the limited video duration.



Fig. 22

The dynamic light and shadow created by falling raindrops also strengthen the intensity of visuals. As the drops hit the ground and splash upward, small highlights appear under the light, adding a sense of motion to the scene and introducing traces of time to an otherwise still space. The changing lights cast on raindrops further the authenticity of the rainy ambience, intensifying the symbolic representation of how historical heritage constantly marches forward with contemporary times.

Upon applying scattered lighting to cultivate a soft setting, the dynamic changes of reflections and light effects of raindrops construct a city image emphasizing the historical inheritance and future-oriented navigations. Light and shadow shape the visual beauty, and connect the past with the present as key visual symbols, making Xi'an City Walls distinctive in the presentation of symbolic significance.

In the short video “I Use a Cinematic Approach to Capture the Old Times of Shanghai”, the creator uses light and shadow symbolically to present both the city’s historical appearance and, more crucially, a sense of personal exploration and introspection within the urban setting. The flow of light and shadow doesn’t just illuminate physical spaces. Additionally, it emerges as a symbolic expression that knits together the city’s visual image with the creator’s inner thinking, resulting in a visually introspective experience (see Fig. 23).

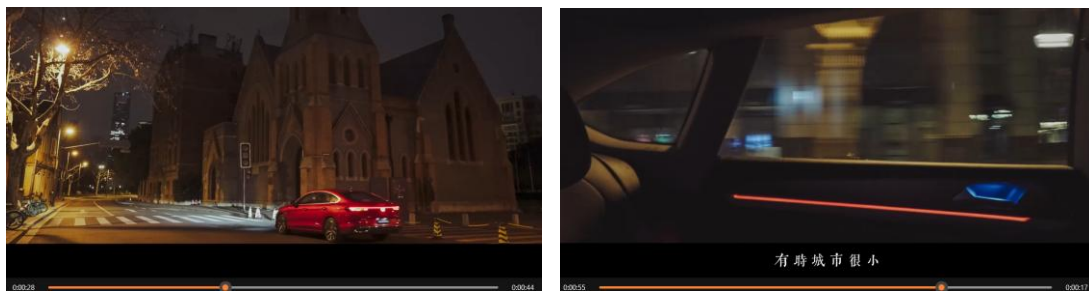


Fig. 23

During rides, vehicle headlights light up the way ahead, integrating themselves with the historical architecture masterpieces along the Bund, and alternating between light and dark effects. Light, as the visual guide during nighttime, signifies the creator’s explorations around the city. Each appearance of light suggests a moment of discovery, turning the city imagery from a concrete display of buildings into a symbolic experience of the urban spirit. Darkness that stands for the unknown, contrasts with the light, which highlights the creator’s curiosity and contemplation, adding a visual tension where exploratory and mysterious identities coexist.

The light and shadow shifting outside the car window further deepen the symbolic meaning of the video. Shots of streetlights, neon signs, and lights from surrounding buildings flicker across the glass, collectively creating a continuous visual stream where the city flashes by as if in glimpses. This constant change of lighting brings a sensation of movement to the frame and functions as a symbol for the life journey. When presenting the city image, the video takes the audience to reflect on the passage

of time and personal experience. The symbolic use of light and shadow turns the city image into more than a representation of physical spaces. Specifically, the city image becomes a medium enabling the expression of personal emotions and philosophical reflections.

The short video builds symbolic connections between explorations, the unknown, and individual experience with play of light and shadow, making Shanghai's urban scenes richer in meaning. The visuals, by moving beyond surface-level presentations, reveal imagery filled with signified significance. In this sense, light both illuminates the cityscape and reflects individual thoughts and exploratory drive.

These shared examples testify to how the symbolic use of light and shadow functions crucially in constructing city images. Unlike functional lighting or mood-setting techniques, this visual approach carries metaphorical meaning, symbolic references, and emotional translation. Creators turn abstract ideas such as the temporal passage, emotional experience, and cultural meaning into specific visual imagery with lighting strategies such as contrast, movement, and composition.

As a result, the video expression unveils an in-depth narrative space that transcends the normal urban presentation. The flowing light and shadow in Beijing's Forbidden City seem to freeze time and evoke a sense of spatial tension. Rain and its reflections in Xi'an's case suggest historical progression. The lights traveling through a car window in Shanghai display inquisitive personal discoveries. These short video works all reveal the symbolic visual power of light and shadow. In all, each city, beyond operating as a physical context, operates as an important vessel for emotions and ideas.

Each city also follows its characteristic path in applying symbolic lighting strategies. In Beijing's video case, the visuals lay greater emphasis on shaping tangible historical

imagery. Light and shadow are used to represent spatial confinement and the temporal flow, reinforcing historical depth and the presence of authority. Xi'an places more stress on the blurred boundary between past and present. Through soft lighting, scattering effects, and reflections, these videos convey each city's cultural depth and layered narrative structure. In contrast, in Shanghai's video, symbolic lighting prioritizes individual experience and mental exploration. The motions of light sources like headlights and neon signs cultivate a sense of bonding between individuals and the city and between individuals and the wider world.

These differences reflect each city's signature historical exclusiveness, cultural context, and visual style, and the interplay of light and shadow to construct city images that are visually rich and imbued with philosophical significance and cultural orientations, lending a stronger intellectual depth and expressive power to short videos.

Emotional symbolism of light and shadow

In short video creation, light and shadow also help communicate the emotional character of a city. According to Grodal (2005), variations in lighting present an emotional tonality to viewers who can thus experience an emotional connection by watching the short video. During city image construction, different approaches to lighting affect the visual presentation, enabling the city image to unravel more enriched meanings apart from functioning as a physical place. The following discussions on city-based case studies dive into the emotional communication of light and shadow in short videos, examining means of adding emotional depth to city imagery with lighting and exploring varied lighting strategies.

In the short video "You Have to Climb the Great Wall at Least Once", light and shadow reflect the emotional sensation reached upon reaching the top of the Great

Wall, weaving a visual language imbued with openness and joy. They are not just part of the visual composition but also the vehicle for emotional expression. This strenuous climbing process is shown as both a physical activity and an intrinsic transformation journey.

In the final part of the video, sunlight spilling over across mountains shapes an illuminous, open visual effect contrasting with the earlier scenes during climbing (see Fig. 24). This shift in lighting reveals the panoramic view from the peak, and symbolizes the climber's mental lucidity and sense of relief. The interplay of light and shadow elevates the climbing action into an emotional breakthrough, expressing in the imagery a vibrating visual and emotional momentum.

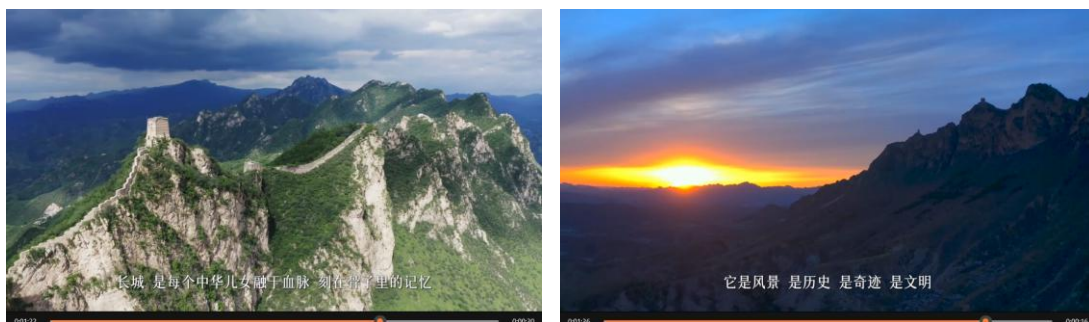


Fig. 24

The video ends on a note of positivity and strength. Such emotional tonality is further reinforced by transitions in lighting amid sunrise, during which natural landscapes take on magnificent looks to symbolize hope and fresh beginnings when the sun breaks through the darkness and climbs above the horizon. By applying light and shadow in a signified way, the climbing is deemed not only a physical challenge but also a navigation of personal growth.

In this way, the video creator successfully conveys the creator's fulfillment and mental clarity upon finishing the climb using light and shadow. Shifts in lighting are

instrumental in emotional representation, capturing the sensations brought by the climbing process in a limited viewing time.

In the short video “I Want to Walk with You from Morning Bell to Evening Drum”, the creator uses nighttime lighting to stress the liveliness of Xi’an’s nightlife. The visuals show an enlivened city scene shaped by light and shadow, and meanwhile, convey the pleasure and excitement of wandering around the night market. The presence of light and shadow pushes the city image beyond physical representation to serve as an instrument for emotional experience (see Fig. 25).

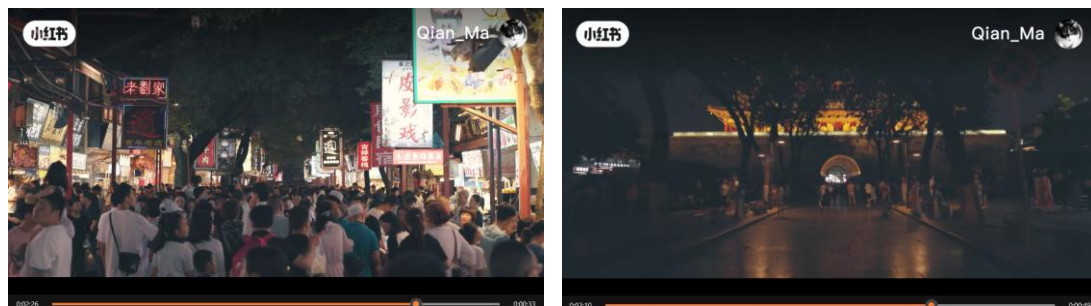


Fig. 25

Glittering lights along the commercial street outline a bright, concentrated visual field under the night sky, enhancing the urban vitality. Flowing neon signs, glowing shopfronts, and the warm-toned lighting from food stalls and storefronts jointly create a lively, rhythmic night market. These lighting effects, while reflecting the prosperity in the area, inject a strong emotional tone to the video, presenting the openness and warmth of Xi’an nightlife to viewers.

Changes in light and shadow further enhance the emotional appeal. Across the night market, flickering lights, shifting colors, and undulating movements of shadow bring diversity and a sense of movement to the visuals. Following an individual’s motions across this place, contrasts in brightness and color changes create ongoing visual stimulation. This seizes the layout details within the night market and reflects the

curiosity and excitement of someone discovering novel experiences. In this way, the lighting helps shape the place as a venue for entertainment and consumption, and turns it into a surprising, joyful location.

Besides simply recording hustling night scenes of a night market in Xi'an, the creator applies light and shadow to construct its ambience, thus presenting lighting-shaped emotional imagery. The lively lights highlight the city's rich energy and activities and simultaneously work as visual symbols to convey the city-specific spirit and the nighttime emotional tone, consequently building a vividly engaging night imagery in Xi'an.

In the short video "What Kind of City Do You Think Shanghai is?", the creator externalizes personal emotions with light and shadow. It presents not only the city's visual features but also an individual's psychological experience in urban landscapes. Shifts in lighting are conducive to shaping the visual style of Shanghai and manifesting the inner sensations of a person. As a result, emotional tones, including the sense of disorientation and distance, are depicted amidst city image construction.

From an overpass perspective that gazes into the distance, nearby low-rises are explicitly portrayed under the light, while skyscraper landmarks such as Shanghai Tower, the Shanghai World Financial Center, and Jin Mao Tower are blurred by lighting differences (see Fig. 26). This lighting treatment shapes a pronounced visual contrast and also implies the creator's uncertainty about what lies ahead. The faraway skyscrapers, softened by light and shadow, seem out of reach in the creator's eyes, since their presence adds a sense of abstraction and estrangement to the city's flourishing evolutions. This establishes the video tonality and visually articulates the creator's mixed emotional state.



Fig. 26

The nighttime landscape along the Bund further reinforces the symbolic significance of light and shadow. The dark silhouette of the creator standing by the river contrasts strikingly against the bright lights of Lujiazui, revealing a sense of distance between the individual and the city (see Fig. 27). Similarly, the liveliness of city lights stands in sharp contrast to the creator's stillness, pursuing emotional expressions through the symbolic role of lighting. The creator appears small and alone under this contrast, which accentuates the grandeur in the city and reflects the isolated mood often occurring to individuals who land in an unfamiliar urban environment.



Fig. 27

Through different uses of light and shadow, the short video presents Shanghai's cityscape as more than a reproduction of physical place by revealing its embodiment of individual emotional experience. As argued by Golden (2015), light and shadow can illustrate the interplay between individuals and their surroundings. In this video case, the creator fully presents her uncertainty about the future and introspection on the sense of belonging in the city, adding a deeper emotional dimension to the city image construction.

The short videos for the three cities show the dual roles of light and shadow in expressing emotions and evoking resonance in urban imagery. Creators shape distinct emotional tones within the visuals by adjusting the level of brightness, color, and direction of lighting, thereby conveying a range of emotions, such as joy, clarity, excitement, or a sense of distance for urban life.

These emotions, instead of being confined to specific narrative events, are communicated through the constructed visual atmosphere, directly striking the viewer's sensory experience. In this way, the city takes on psychological-level

functions that transcend its physical presence. Light and shadow have emerged as a subtle emotional language infusing the city image with warmth and emotional impact. To illustrate, the viewer can sense the radiance of sunlight and satisfaction upon climbing to the top of the Great Wall in Beijing, experience the sustained flows of energy and cadence brought to life by night market lights in Xi'an, and the loneliness of dwelling in the metropolis of Shanghai as reflected in the mix of neon lights and shadow.

Although all three cities apply light and shadow to emotion representation, they vary in lighting approaches and emotional focuses. In Beijing's video case, the emotion representation is more contingent upon natural light. Elements such as sunrise, mountains, and open spaces create a vibrating rhythm of emotional relief and renewal, reflecting a shift toward a more optimistic mindset. Xi'an's video emphasizes the vibrancy and warmth in nightlife, where neon lights and commercial settings convey openness, friendliness, and everyday vitality. The video depicting Shanghai, in contrast, takes a more introspective, personal approach. Contrasts in distance and brightness used by creators all lead to the representation of estrangement and self-identity issues within a bustling urban setting.

While short videos of all three cities express emotions with light and shadow as an effective medium, their tones and types of emotions, and how they guide the viewer's psychological experience differ. These differences reveal that emotion-based lighting strategies employed in short videos adapt to certain city contexts and suggest an extensive range of expressive possibilities.

Conclusion

With the continuous evolution of media expressions, the roles of light and shadow in short videos are continuously rising in prominence. As showcased by Säks (2023),

light and shadow not just provide decorative elements to enhance visual appeal, but also stand as the key visual strategy for shaping the ambiance, conveying intentions, and helping the audience grasp the city's image. This chapter, grounded on this argument, concludes that varied design and use of light and shadow both support visual composition and undertake multiple functions, including cultural translation, emotional expression, and construction of impressions within a limited time.

Today's media context, led by the mobile viewing experience, sees how light and shadow bear intricate information and sensory cues in highly efficient and condensed formats. Light and shadow draw the audience to swiftly recognize the multi-layered city characters and forge emotional connections with cities, going beyond creating visual experiences to set up a mode of expression in between imagery and storytelling. It links the city's physical functions to its cultural meanings and activates a sense of psychological resonance. As a result, representation of cities in short videos holds stronger expressive power and influences at both emotional and symbolic levels.

The variations in lighting strategies across the video cases of Beijing, Xi'an, and Shanghai manifest creators' responses to the distinct character of each city.

In Beijing's video cases, light and shadow are applied to highlight historical depth and the authority of spatial structure through stark contrasts and orderly composition, resulting in a dignified city image. In Xi'an's video cases, lighting, as a medium for cultural expression, finds its applications in evoking traditional imagery in the viewer's mind and shaping the rhythm of everyday scenes such as night markets. This thus shows a blend of cultural heritage and contemporary vitality. Shanghai's video cases place greater emphasis on the fluidity and variety of lighting, where imagery, such as lights flashing through car windows, shifting neon colors, and towers lit up at

night, condenses the fast-paced, sensorily intense, and emotionally complex dimensions of modern urban life.

In brief, the diversity of urban styles and creators' adaptive usage of lighting techniques, as per each city's characteristics, are presented in these differentiated lighting strategies. Condensed layers of meaning are unfolded in these videos of short duration and limited spaces. Such strategic treatments of lighting, in addition to showcasing creators' observations within each city, unveil the potential of light and shadow in constructing city images under the new media context of short videos.

Chapter 5 The Role of Sound in Constructing City Images

Introduction

Sound, as another crucial element for city image construction in short videos, works in conjunction with visual elements to present the city's space, cultural atmosphere, and emotional tones. Despite the leading role of visuals and the supporting function of sound in video production, sound reinforces the viewers' perception, conveys emotions, and helps with content structuring. These benefits are especially crucial for short videos characterized by a limited duration, fast narrative pacing, and extensive editing. In brief, sound is key to the representational ways of cities, facilitating the connection and transition of scenes, contextual information, and guidance for viewers.

Background music, ambient sound effects, and voice-over operate as the three most representative elements in the sound system of short videos, each of which contributes to city image construction from a different perspective. This chapter comparatively studies video cases of Shanghai, Xi'an, and Beijing, exploring the respective functions and applications of these three elements in cities' image construction. It aims to reveal the roles of sound as a representational technique in shaping diverse city imagery, thereby suggesting a new perspective for comprehending the variety of urban expressions in short videos.

Background music in city image construction

Background music stands as the most emotionally expressive among sound components in short videos, and is central to shaping the emotional ambience of a city. Compared to the authenticity of sound effects and the explanatory nature of voice-over, background music prioritizes mood establishment, regulating the narrative cadence, and evoking emotional resonance in viewers.

In short videos characterized by condensed visuals yet dense information, it helps build a subtle auditory structure through melodies, rhythms, and emotional progression. This translates the city's warm vibe, rhythmic sense, and spatial experience into perceivable auditory sensations, through which viewers emotionally relate to content in a short time (Berland, 1993). Such emotional engagement thus constructs city images step by step.

The category and style of background music often elicit cultural relatability in viewers. For instance, rhythms, instrumentation, and melodic patterns of varied musical styles can be associated with cultural attributes, historical memories, or the city's atmosphere. Hence, it functions as a symbolic cultural element for city image construction (Duan, 2022). It also helps with emotional expression and facilitates the visual translation of cultural identity and atmosphere reconstruction, emerging as a key vehicle that bridges city images and viewers' perception.

This section discusses the employment of background music in short videos in the lens of emotional construction and cultural expression, as well as the musical strategies applied by creators in image construction.

Background music in emotional construction

In short videos, background music not only sets the emotional tone for visuals but also subtly shapes the perception and responses from viewers to the city's overall character. This section dives into its usage in forging the emotional ambience across cities of varying features. The three city case studies inquire into how creators coordinate music with visual rhythm to stimulate emotional resonance with the city's image. In this process, music, aside from being the "soundtrack" for the visuals, highlights the complementary function of sound during city image construction to

drive this construction process.

In Beijing's short video "In *This Winter, You Have to See the Golden Light at the Seventeen-Arch Bridge in the Summer Palace*", a Chinese male singer Zhou Shen's song *Flickering Light* is set as the background music. The masterfully coordinated design of sound and visuals constructs a city image rich in historical depth and cultural empathy. This musical strategy, beyond presenting its decorative function, acts as the medium for reconstructing and providing emotional guidance to the city space and its cultural meaning.

In imagery representation, music builds a strong semantic connection with the natural landscape. The "Golden Light Piercing Hole", a rare phenomenon around the Winter Solstice in the Summer Palace, appears when golden sunlight pierces through arches of the Seventeen-Arch Bridge, and bridge reflections ripple over the water. This spectacle delivers a striking, timeless scene of the imperial garden (see Fig. 1), whose glittering movements of light and shadow are depicted by the airy, distant melodies from the background music, *Flickering Light*. This music-enabled sound strategy aligns with the visuals to produce a "synesthetic simultaneity" that enhances both immersion and visual aesthetics, thus reinforcing the cultural memory deep-seated in the natural cityscapes through symbolic resonance.



Fig. 1

In terms of emotional expression, this piece of background music arouses a mood that seamlessly echoes with the theme of the video. Its lyrics articulate a poetic dialogue between the past and the present: *“Your traces travel through thousands of years; you said the moonlight once shone upon the same city wall.”* These lines metaphorically link the continuity of time with the permanence of the city, evoking modern reverence for historical relics. The music’s gentle tempo and flowing melody further reinforce this sense of temporal fluidity, aligning with the short video’s portrayal of the moment when scenes witnessed by modern people are the very ones once beheld by those who lived in ancient times. Through this lyrical and sonic interplay, the composition amplifies the time-transcending visual experience and deepens the audience’s emotional connection to Beijing’s layered history.

The background music creates an emotionally impactful ambience, sublimating the historical landscape from a visual component to a compelling medium that cultivates in viewers a shared sense of cultural identity. As inferred from DeNora (2000), music is not only a product of emotion but also an instrument that orchestrates the presentation of feelings within social contexts. Apart from admiring the scenery,

viewers can be empowered to feel as if the city awakened its historical memory. Hence, this sound strategy is designed to strengthen emotional immersion and emphasize the continuity of Beijing's historical identity.

The background music in Beijing's video case makes both symbolic and emotional contributions, constructing a cultural and affectionate connection that transcends time and space. This approach presents the creator's active use of sound aesthetics during city image construction in the short video era, and the layered connections between the city's rich historical texture and its emotional resonance mechanism in today's media context.

In Xi'an's short video "*To Long for Each Other in Chang'an*", shifting musical rhythms and a blend of styles portray Xi'an's emotional progression from nighttime stillness to daytime vibrancy. The sound design, in synchronization with the cadence of visuals, effectively guides viewers to emotionally perceive the changing urban atmosphere (Pavlović, 2011).

The video also records a clear contrast and gradual build-up between the music rhythm and the city's nightlife vibe. The background music in the hotel lobby unveils the video, featuring a slow-paced melody typical of ancient towns, and calm, steady tunes that constitute an auditory environment with lower dynamism. The low sound pressure and restrained mood naturally match the tranquil space, and set the tone for urban nightfall. This moment witnesses how music fills the acoustic void within the environment and leads viewers to immerse themselves in a peaceful and progressively easing frame of mind. This emotional setting prepares for shifts in rhythms that follow and clarifies structural transitions in pacing.

When the camera shot shifts to the central area of Xi'an's Grand Tang Shopping Mall, the musical rhythm noticeably changes. A brief, powerful drumbeat, as a sonic cue,

signals the official start of the city's ever-bright nightlife and instantly activates the viewer's emotions. Traditional instruments such as the bamboo flute, ruan, and pipa are then introduced in succession (see Fig. 2). The melody moves into a pentatonic mode characteristic of traditional Chinese music, and as the plucked timbres and added percussion layer in, the texture thickens and the tempo rises. The meter alternates between duple and triple groupings, producing a sense of forward momentum that matches the increasing density of lights and crowds. This rhythmic escalation functions as a sonic metaphor for the city's shift from early-evening calm to lively nocturnal energy.



Fig. 2

The polyrhythms-enabled layered ensemble sensorily expands the soundscape, which is closely coordinated with the constantly shifting visuals, including lights, performances, crowds, and street scenes. This amplifies the synchronization between sight, sound, and emotion. Music applied at this point is no longer just a background element, but also the key driver behind viewers' immersive experiences (Baker et al., 2016). This alignment between musical rhythm and bodily movement is considered

sound's fundamental way of guiding everyday behavior and shaping the surrounding social atmosphere.

In addition, the creator skillfully introduces the prelude from the traditional-style song *Free Immortal* near the video's end, further enhancing the emotional tone and rhythmic variation. This more loosely structured melodic phrase stands in gentle contrast to the dense, ensemble rhythm that previously occurred. It effectively softens the intensity of earlier sounds and serves as a transitional passage, helping to ease the emotional buildup and bring closure to the night scene's expressive charms.

The creator's music strategy displays a powerful grip over rhythmic shifts and emotional layering. It enables the background music to flow in sync with the visuals, thanks to the smooth transition from mellow musical performances in serene scenes to ensembles in lively settings. This coordinated progression emotionally portrays the gradual unfurling of Xi'an's nightscape.

In the short video "*Ten Years Later, I Still Miss the Shanghai Depicted in Micro Moments*", background music voices emotions and, more crucially, threads through the video's emotional arc, contributing to the holistic ambience construction of the city. The theme song from the film *Tiny Times* is taken as the soundtrack for this short video, bridging image and memory via music design in narrative and sensory dimensions.

The video begins with the visuals that depict city glimpses in the creator's memories, paired with a low-pitched, restrained melody from the song's opening section. The music applied here features a slow tempo, closely woven harmonies, and a relatively unvaried melodic line, creating a subdued, slightly heavy emotional tone. The visuals, including old alleys, a grey-blue sky, and the creator's slow-paced walk, align with music to shape the city image as nostalgic and slightly melancholic (see Fig. 3). At

this stage, sound, besides its background role, also guides the viewer's emotional engagement, making the city visually perceivable and emotionally tangible.



Fig. 3

Notably, as the short video draws to an end, the musical tone shifts to an abundantly hopeful, warm register. This shift harmonizes with the lyrics' expression of enthusiastic outlooks on the future, and presents the creator's emotional transition from introspective reflection to forward-looking anticipation. The visuals capture scenes where the creator gazes over the sky at night and gradually merges herself into the crowd, actions that align closely with the melody's rising emotional tides (see Fig. 4).

The synchronization between music and visual rhythm creates a city perception that unfolds in an inside-out manner. It begins with personal emotion, expands into the urban space, and ultimately circles back to voice hope for the city's future and individuals' aspirations. In this particular moment, the blend of music and city cultivates a mutually reflective and amplifying emotional space.



Fig. 4

Overall, the video effectively integrates personal memory, urban space, emotional recollection, and real-life street scenes by structurally referencing and emotionally matching the film's theme song (Boltz et al., 1991). Through shifts in rhythm and emotion across different periods, music works in tandem with visuals to unfold the city's emotional ambience, thereby enhancing the overall audiovisual experience. In this process, background music guides emotional identification, helping viewers psychologically relate to the scenes through its melody and rhythm arrangements. The creator, by employing the narrative thread of music, builds an "emotional version of Shanghai" where memory and hope intertwine, constructing a more expressive, layered auditory presence for the city image.

The three city cases show that creators differ in their sound strategies concerning music choices and rhythmic arrangement. These strategies not only support the narrative but are also instrumental in constructing the city's emotional representation, as shown in the visuals. Music, as an emotional driver for short videos, helps shape viewers' impressions and guide their emotional interactions. Despite its shared function across the three cities, its approach to emotional tone and expression varies

across cities.

Beijing's video case emphasizes solemnity and calamity in its music strategy. The application of lyrical and emotionally resonant melodies enables the sounds to echo the historical spaces featured in the video, establishing a tranquil yet ceremonial mood and adding a strong sense of historical consciousness to the city's image. In contrast, in Xi'an's case, its sound design is more focused on rhythmic flow to portray the city's emotional shift from muted to lively at night through the gradually ascending buildup of pacing. This highlights Xi'an's openness, energy, and extensive public engagement. In Shanghai's video case, the approach to background music draws on film music. It crafts a fluid emotional path that starts from personal memory, then travels through city space, and reaches the emotional uplift. This nurtures a strong resonance between individual introspection and urban imagery, resulting in a nostalgic and hope-infused city experience.

The sound strategies in these three city cases distinguish in the musical narrative rationale they apply: "historical immersion" for Beijing; "rhythmic transition" for Xi'an; and "emotional projection" for Shanghai. These case studies collectively reveal that music undertakes more than the supporting function of constructing the city's visual image in short videos. They also present the power of music in emotion arousal, atmosphere building, and enhancement of audience perception. By shaping rhythm, tone, and instrumental texture, music forms a dynamic connection with the city space. This interaction allows viewers to quickly become engrossed in a specific emotional context and thus, more effectively articulate the city's emotional identity within constrained viewing time.

Background music in cultural construction

Beyond its emotional enhancement effect, background music also plays a significant role in cultural expression. This subsection delves into its function in urban culture representation, covering comparative analyses on the application of background music in short videos featuring the three city cases as illustrated above. It explores how creators employ traditional folk melodies, musical instruments of ethnic groups, or modern pop elements to construct culturally distinctive impressions. Consequently, this approach strengthens the cultural symbolic significance of these cities in media dissemination.

In the Beijing-oriented short video “Beneath the Great Wall, Intangible Heritage Shines Through Time”, the applied background music not only enhances the audiovisual unity but also reflects the creator’s purpose of embodying the city’s historical and cultural identity through musical style and mood. The video centers on two pieces of cultural heritage, namely the Great Wall as a UNESCO World Heritage site and the Nanjing velvet flower craft as an intangible heritage. Jay Chou’s *Lan Ting Xu* as the soundtrack draws on the unique atmosphere and emotional tone of traditional-style music to construct a rich cultural image of the city, as seen from the video.

Regarding musical appreciation, *Lanting Xu* blends music of guqin and flute and chant-like melodies to create a peaceful, elegant, and classically inspired ambience. This musical style, in close alignment with the historical and cultural themes presented in the video, operates music as a medium that deepens spatial memory and traditional values beyond supporting the scene as a background tool. The use of traditional-style music in sound strategy reflects the creator’s cultural awareness. Specifically, it leads the audience into a particular historical setting through musical expressions, strengthening the cultural recognizability of Beijing’s city image.

In terms of lyrical content, *Lanting Xu* expresses admiration and reverence for the passage of history, whose emotional tones are echoed and reinforced in both narratives and visuals of the video. By depicting the grandeur of the Great Wall and the delicate craftsmanship of Nanjing velvet flowers, the creator harmonizes the rhythm of the spoken words and the emotional delivery with the acclaim for traditional Chinese culture embedded in the lyrics. This consistency across lyric, narrative, and emotional tone turns music into an integral part of the city's cultural expression, deepening the audience's relatability with the video's theme.

In addition, the intimate interplay between the musical rhythm and the video's narrative structure is identified in the study. The visuals focus on the natural scenery and historical landmarks along the Great Wall during the musical prelude, crafting a visual tone rife with calmness and timelessness. At the onset of lyrics, the video transitions to a featured, examined gaze at the velvet flower craft, highlighting the well-recognized craftsmanship and the cultural heritage value (see Fig. 5).

As music builds up to its emotional culmination, the visuals take another shift to the creator's reflections on Beijing's cultural identity. This layered coordination between musical rhythm, visual content, and emotional structure ensures smooth emotional progression, unraveling the city image that boasts a distinctly paced and emotionally rich expression.

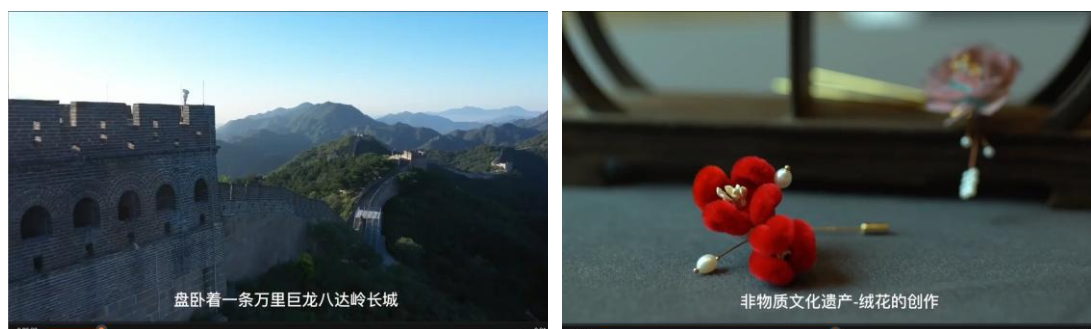


Fig. 5

This short video constructs the city image through the strong match of traditional-style background music and its cultural theme. The multidimensional synergy of melodies, lyrics, and narrative rhythms shapes the urban representation. In this context, background music goes beyond simply establishing the emotional tonality by emerging as key to expressing the city's historical and cultural identity. Hence, it lends a more profound, distinct cultural footprint to Beijing's image on an auditory level.

In the short video "*Under Xi'an's City Wall Runs the Train of Xi'an People*", its background music even transcends the function of emotional enhancement, becoming a central element tightly bound to the city's cultural identity. The video features the local pop song *Song of Xi'an People* as its soundtrack. This musical piece precisely grasps regional identity and emotional expression of Xi'an residents through its melodies, lyrics, and tones, which makes the city culture more tangible, emotionally expressive, and collectively memorable.

Song of Xi'an People boasts a layered musical interpretation. As a product of local popular culture itself, its simple melodies, steady rhythms, and performance in local dialects all cultivate a vivid sense of location via sound design. The lyrics, centered on everyday life in Xi'an, interweave themes of local transportation and historical landmarks. Lines like "The train passing beneath city walls belongs to Xi'an people." combine space with identity, allowing listeners to experience the daily rhythms and cultural atmosphere in Xi'an (see Fig. 6). In this context, music, besides being an emotional vessel, distills the city's cultural experience and local spiritual essence.



Fig. 6

The video's visual composition synchronizes with the music progression, forging a narrative structure where sound takes the lead and visuals complement it. As the song opens, the visuals present close-up shots along city walls, visually guiding the viewer into Xi'an's historical space. In the face of lyrical mentions like "train" and "streets", scenes of subways in motion, commuters, and lively streets that present a modern view of the city emerge in the visuals. This sound-image coordination naturally blurs the line between tradition and modernity. Notably, the visuals do not overpower the music, but respond to such lyric-shaped city imagery, reinforcing the leading role of sound in cultural construction.

In addition, toward the video's end, the creator introduces a harmony section with slightly rhythmic variations, while the visuals shift to more emotionally rich content, such as family moments, facial expressions of citizens, and sunset cityscapes. This overlay of "musical structure and visual emotion" drives the narrative from a cultural urban expression toward a cultural identity rooted in personal memory. As a result, viewers embrace shared emotions by listening to this video. This layered tripartite

structure of “music–scene–identity” enables Xi’an to be perceived as both a vibrating physical space and a city with a sound-shaped distinct character.

Applying local pop song and a sound-first storytelling approach to Xi’an’s video case moves beyond defining city culture through the sole medium of visuals by distilling, understanding, and sharing the culture with sounds. In this context, background music bolsters the representation of cultural identity, imparting a recognizable texture and auditory emotional warmth to Xi’an’s city image.

In the short video “*Walking Through Wong Kar-wai’s Shanghai*”, the creator thoughtfully selects *Yumeji’s Theme*, the iconic soundtrack from the film *In the Mood for Love*, as the background music. The piece itself doesn’t originate from a cultural context in Shanghai, but is widely recognized as a signature sound of director Wong Kar-wai’s cinematic style. It symbolizes restrained emotions and a poetic sense of distance often associated with Eastern urban settings. Its ingenious incorporation into a short video elucidating Shanghai’s imagery evokes a familiar, nostalgic audiovisual tone and subtly draws a cultural connection between Shanghai and Hong Kong through sonic metaphor.

Yumeji’s Theme repeats the melodic motif primarily played by the viola to render a sense of circularity, restraint, and introspection. Such melodic structure carries a distinct emotional quality of “delayed unfolding”, well aligning with the slow rhythms of scenes in the video—old streets, narrow alleys, local walkways, and passersby—infusing the entire piece with a clear “Wong Kar-wai’s atmospheric style” (see Fig. 7). The delicate emotions in the music echo Shanghai’s visual details, sketching a tranquil, vague cultural portrait of Shanghai that contrasts with the conventional image of this fast-paced modern metropolis.



Fig. 7

The soundtrack choice in this short video brings forth a cross-city cultural association. *In the Mood for Love* portrays Hong Kong in the nostalgia-filled 1960s, when individuals were compellingly connected with each other. The incorporation of its signature theme song into the video case of Shanghai prompts viewers to identify an urban atmosphere and emotional register similar to the film. As a result, Shanghai is presented as an Eastern and modern label seamlessly blending soft and aloof qualities. This sound strategy not only adds a unique narrative rhythm to the city space but also strengthens the video's cultural expression through its evocative and relatable power.

In addition, music plays a central role in setting the emotional tone and shaping the video's narrative pacing. The short video avoids using voice-over and rapid editing, and instead features a repetition and continuity of the soundtrack to keep the rhythmic steadiness. The slow-paced, reflective viewing experience it generates draws the audience into contemplative gazes of the cityscape. This music-driven approach to city presentation doesn't pursue information density but emphasizes the buildup of immersive atmosphere and emotional resonance. As a result, it nurtures a city culture characterized by an ineffable yet profoundly sensed mood (Bezdek et al., 2022).

Overall, this short video builds a refined, slow-paced, and metaphor-rich acoustic environment by adding *Yumeji's Theme* as the film soundtrack. This approach endows Shanghai's city image with a sense of cultural extension transcending its geographic boundaries (Yeh et al., 2008). The background music evokes cultural associations for viewers and constructs a clearly recognizable urban portrait of Shanghai.

Short video cases of Beijing, Xi'an, and Shanghai distinctly vary in their modalities of cultural expressions when selecting and applying background music. While all three cities use music in city image construction, the specific strategies differ significantly based on cultural backgrounds and communication objectives.

Beijing's video case tends to favor music with classical aesthetics or traditional undertones. The interplay of steady melodies, restrained structure, and lyrical connotations in the applied background music underscores the city's historical profundity and cultural dignity. This approach crafts an auditory atmosphere that aligns with the city's heritage, allowing viewers to perceive a ritualistic sense and cultural depth that reflects this Chinese capital city's rich identity.

Xi'an's video case places greater emphasis on the connection between music and local lifestyles, whose musical choices rely on straightforward melodies and emotionally grounded rhythms, presenting a warmhearted, everyday cultural impression. This strategy highlights the harmony between sound and space, reinforcing the link between regional identity and daily urban experience. As a result, the city's culture comes across as more relatable and emotionally accessible.

Shanghai's video case gears toward a more stylized, aesthetic approach to background music, focusing on the harmony between musical atmosphere and urban character. The strategy emphasizes the suggestive and associative nature of sound, using restrained rhythms and lingering melodies to convey a sense of elegance, complexity, and cultural tension within the city.

The three cities are varied in their cultural expression strategies through background music: "adherence to tradition and ceremonial practices" in Beijing's video; "embodying everyday details and rich in historical heritage" in Xi'an's video; and

“aesthetic and atmospheric imagery” in Shanghai’s video. These discrepancies reveal the diversified applications of music by short video creators to present emotions and more profound implications. Music helps articulate each city’s cultural positioning, communication style, and insights into the audience’s aspirations. In this way, background music acts as a bridge between cultural content and media perception in cities, making city culture more tangible and emotionally resonant through sound-based practices.

Sound effects in city image construction

In short videos, sound effects function as a highly authentic audio element key to constructing the authenticity and sensory credibility of city spaces. Background music further accentuates the emotional tones and cultural symbolism, while sound effects are closely tied to the acoustic foundation of urban life.

In typical terms, instead of appealing to the audience through melodic presentation, sound effects evoke a sense of immersion and presence through their everyday and immediate nature. This helps elevate the viewing experience from mere visual aesthetics into a more textured, multi-sensory perception (LaBelle, 2019). This sound-driven place no longer casts the city into an image to be “seen”, but a space to be “heard”. Viewers, when presented with familiar sounds in short videos, begin to recognize the city rhythms, envision its atmosphere, and sense its social dynamics.

Moreover, the roles of sound effects in short videos go beyond recreating the environment. Creators are found to purposefully choose sounds and apply sound effects as symbolic and emotionally charged elements, turning sounds from simple real-world representations into tools for emotional guidance, cultural association, or narrative transition. These sounds can evoke sensory impressions of a place, assist

with the building of mood, reinforce emotional tones, or even serve as metaphoric expressions of a city's character or cultural identity.

In this sense, sound effects not only enrich the city's sensory texture but also strengthen its cultural construction through sound design (Gou, 2025). This symbolic transformation engages sounds in the meaning construction path of city representation, turning transient auditory elements into culturally significant and emotionally resonant markers.

Sound effects in city authenticity representation

In short video production, sound effects are not just used to recreate the environment. Their familiarity, continuity, and sense of spatial presence help construct a concrete, conceivable, and alluring city image. Through comparative research on the sound effect strategies spanning the three cities' video cases, this section investigates how sounds are used to enhance a city's recognizability, add depth to spatial expression, and activate context-specific memories associated with the city.

In the short video "*Beijing: 'Bei' of Beiping, the 'Jing' of Capital*", sound design sits at the core of forging urban authenticity. This video selectively features environmental sounds holding geographic and everyday implications, instead of adding background music to convey emotional atmosphere. The applied sounds quickly lead the viewer into an auditory setting strong in its sense of place, and evoke shared memories of viewers, turning the city from a merely perceivable object into an entity to be "heard and lived".

The video opens with the crisp sound of a rickshaw bell, immediately establishing the history-shaped soundscape. Rickshaw, once a transportation tool rooted in historical associations, has vibrantly emerged as an icon of traditional Beijing's street culture.

Its clear and rhythmic bell sound instantly transports viewers into the cultural space of Beijing's hutongs at the audiovisual level (see Fig. 8).

Beyond reproducing a physical reality, its sound symbolically functions by inviting the audience to imagine the daily life infused with traditional rhythms and spatial orders from the past, constructing an auditory image in Beijing hutongs. Its pronounced geographical identity transforms it into a sensory trigger purposefully embedded by the creator into short video production. Its familiar tone evokes emotional resonance and draws viewers into the context constructed within the video.



Fig. 8

In addition to the rickshaw bell, the video features two other sound effects that capture everyday life in Beijing: birdsong heard as bird lovers talk with their pet birds in public parks, and the rhythmic sound of shoes striking the ground during shuttlecock play. These ordinary sounds frequently appear in specific places such as Beihai Park or near the Temple of Heaven, and grow into part of the so-called “Beijing’s daily soundtrack”.

The bright and intermittent birdsong is commonly associated with the lives of elderly park-goers, while the sound of shuttlecock play reflects a relaxed, open, and socially engaging lifestyle (see Fig. 9). Through purposive infusion of these everyday sounds into the visuals, the creator stirs up auditory memories of life in Beijing through selective sound details. This approach helps construct a city image both authentic and emotionally intimate to the viewer.

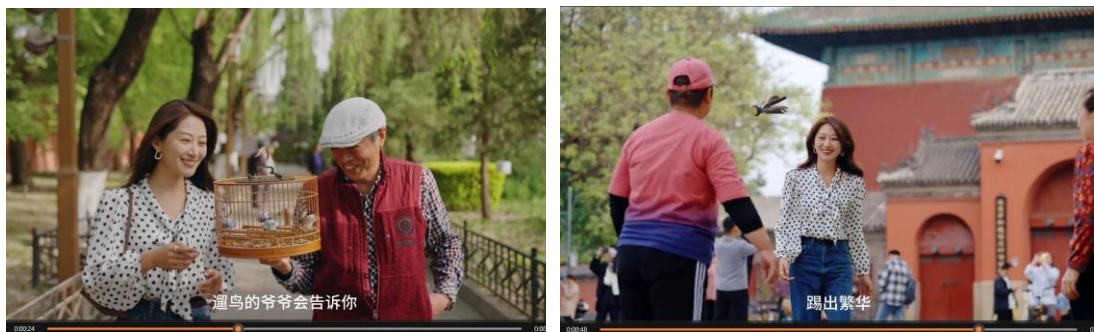


Fig. 9

In this case, the creator combines and highlights sound effects that carry both spatial symbolism and daily feels, effectively reconstructing the city’s “audibility”. Sound doesn’t just complement visuals but also evokes memory, stimulates imagination, and further defines the city’s identity. This approach injects a richly detailed and auditorily authentic spatial texture into Beijing’s city image construction.

In the short video “*Life Advice: You Must Experience Xi’an’s Morning Market*”, sound effects, besides supplementing the visuals, are key to arousing the everyday rhythms and enhancing the sense of spatial authenticity. By carefully selecting and amplifying ambient sounds arising from food preparation and stall operation, the creator constructs a lively, immersive morning scene. Beyond “seeing” the hustles and bustles of Xi’an’s morning market, viewers are attracted to “hear” and “feel” its rhythms, warmth, and everyday atmosphere.

Audibly recorded detailed moments surrounding food preparation bring forth the most striking sounds in the short video, which encompass the crisp sound of meat pies breaking fresh out of the oven, sizzling sound of food fried with hot oil, and the rhythmic clattering sounds of chopping, flipping, and scraping at food stalls (see Fig. 10). Instead of being incidental background noises, they have been intentionally recorded and emphasized to heighten the impactful food imagery. The presence of these food-related sounds deepens the viewer’s sensory immersion and sparks “taste-based associations” that facilitate an indirect but compelling experience of Xi’an’s culinary culture. At an auditory level, this approach cultivates Xi’an’s identity as a “city of everyday life” brought to life by sound effects, allowing city culture to resonate emotionally through the sounds of street food.



Fig. 10

In addition, sound effects in this short video highlight the finer details of city life and enhance the spatial authenticity. The morning market in Xi’an is naturally a sound-rich and layered environment. Nonetheless, the fast-paced, scroll-based viewing format of short videos leads to the frequent negligence of countless subtle yet culturally meaningful sounds. Consequently, the creator intentionally captures and

amplifies details such as the murmurs in conversations from the crowd, rustles of plastic bags, tinkles of coins, and the casual dialogue between customers and vendors. These sounds all enable the audience to “hear” the easy-to-overlook layers of urban life.

By emphasizing these small auditory details, the video increases the scene’s sense of authenticity and reveals the simplicity, warmth, and rhythmic vibrancy of everyday life in Xi’an. This focused portrayal of sounds helps strengthen the emotional links between sounds and the city’s daily routine.

Through its dedication to food-related and ambient sounds in daily life, the video constructs an invigorating and atmospheric morning scene in Xi’an. Sound effects become crucial to engaging the audience’s sensory experience, investing the urban life with a tangible, textured sound-shaped presence. This again converts Xi’an’s city image into an exuberant experience that can be “seen” and “heard”, or more specifically, a warm, living soundscape.

In Shanghai’s short video “*Has Autumn Reached Your City?*”, the creator carefully captures and arranges natural sound effects to vividly present the autumn ambience in the city. Unlike background music highlighting emotional tones, sound effects help express seasonal change, triggering physical sensation and enhancing the spatial authenticity. Through sound, Shanghai, this highly urbanized city, is endowed with a deeper emotional layer shaped by the passage of time.

The soft crunches of footsteps on fallen leaves along city streets are most prominent among the sound elements in the video. Being light, intermittent, and slow in rhythm, it effectively evokes the stillness unique to autumn while activating the viewer’s bodily memory of seasonal changes. The sound, when especially paired with visuals of quiet streets at dusk and slow-moving footsteps, functions as an auditory symbol

depicting the city’s slowing pace (see Fig. 11), and leads viewers to “hear” autumn in the city. This kind of sound design strengthens the temporal perception within the urban space, turning the otherwise abstract experience of “engaging with autumn” into concrete, recognizable experiences through sounds.



Fig. 11

The video also portrays a sound effect in the suburban scenes to introduce a sense of spatial depth, where the creator applies the soft rustle of wind passing through withered grass. Unlike the footsteps on city streets, this natural ambient sound expresses openness and air movement, cultivating an “open and spacious” soundscape. Visually, it corresponds with empty lots, quiet roads, and low-density buildings on the urban outskirts. The wind sound connects the visual experience to soundscapes, reinforcing transitions between the city and the edges of natural landscapes (see Fig. 12). This sound not only enhances the environment’s authenticity but also introduces an emotional tone of seasonal distance and restraint. It allows viewers to perceive Shanghai as a “lived-in place” brimming with natural elements and intrinsic urban rhythms beyond considering it a hustling “concrete jungle”.



Fig. 12

Overall, this short video employs carefully selected ambient sounds, such as footsteps and wind movements, to audibly reproduce the autumn ambience in Shanghai. By using sound effects as key cues, the creator seamlessly links the city's physical space with seasonal changes, daily rhythms, and the passage of time. This approach allows viewers to engage with the city through this auditory feast and acquire a more authentic perception of its atmosphere. Sound, besides shaping the texture of physical space, arouses subtle experiences associated with everyday emotions, changing weather conditions, and inner rhythms of the city. This sound strategy enriches Shanghai's identity as an urban space, thereby presenting it as a multi-layered sensory environment.

The comparative analysis of the short video cases from Beijing, Xi'an, and Shanghai demonstrates that sound effects, as a crucial medium for restoring urban authenticity, serve constructive yet varying roles in each city's imagery representation. Beijing's video case places greater emphasis on historically emblematic sounds, such as rickshaw bells and hutong ambience, to elicit cultural memory, building the perception of authenticity on recognizability and continuity.

Xi'an's video case focuses on detailed, sensory-rich sounds from its morning market, such as the sizzles of food, vendor movements, and human conversations, which closely align with the visuals, highlighting the city's warmth and everyday vitality. In contrast, Shanghai's case applies natural ambient sounds, like the crunch of fallen leaves and the rustle of dry grass over the wind, reflecting the city's changing rhythms and seasonal shifts. These enriched sound elements together build the auditory-perceived time and space and cultivate a restrained, calm, and rhythmically textured urban atmosphere.

As seen from the comparisons, the short video for each city shows a different priority on sound usage: historical continuity for Beijing; intensity of daily life for Xi'an; and mood and rhythm development for Shanghai. These videos all prove that sound effects in short videos undertake multiple roles by standing as background elements and being instrumental in shaping the city's auditory experience. Sound enables viewers to move beyond visual observation towards engaging with the city through immersion in the soundscape, arousing in them a deeper recognition of urban identity, spatial texture, and everyday emotions. Short videos enhance the urban authenticity and accessibility of local cultural experiences thanks to the strategic integration of these sound elements. In this way, sound effects become an essential technical tool in city image construction.

Symbolic meanings of specific sound effects

Beyond the authenticity cultivation, sound effects in short videos are often infused with symbolic functions, bearing emotions, suggesting spatial meanings, and shifting the significance of various narratives. This section focuses on the symbolic roles of specific sound effects in city image construction. By analyzing representative examples from short videos filmed in Beijing, Xi'an, and Shanghai, it explores how

creators apply selective sound design to extract certain sounds from their original contexts and turn them into narrative cues and cultural symbols. In this sense, the section further reveals the power of sound effects in shaping city imagery within the short video format.

Beijing's short video "*Only a Snowy Hutong Holds the Warmest Human Touch*" doesn't define the winter coldness as the dominant emotional tone in the visuals. Instead, its creator has meticulously selected and arranged specific sound effects to build a warm, intimate auditory context. Despite its visual focus on scenes typical of a winter day, ranging from snowy streets to pedestrians in a hurry, the emotional tone doesn't convey a sense of distance or iciness. Rather, its gentle warmth is presented by the blend of motion and stillness, exterior and interior, to jointly build a sensory experience of "winter comfort."

The addition of lively indoor sounds layered over outdoor street scenes (see Fig.13) operates as one of the most symbolically meaningful choices. These sounds don't originate from the original visual environment but are intentionally inserted by the creator to guide viewers toward a perception of life inside spaces surrounding the city. Layered sounds include the conversation murmur, clink of dishes, and imagined sound of rising steam, all of which contrast with the cool-toned visuals and bring about a sense of warmth hidden behind closed doors.

This contrast invites the audience to imagine the daily coziness tucked away inside the hutongs. In this way, the city is portrayed with an "outwardly cool but inwardly warm" emotional structure. Sound effects also build up a bridge between opposing feelings, so that the audience can perceive traces of human warmth even across the snowy, silent streets.

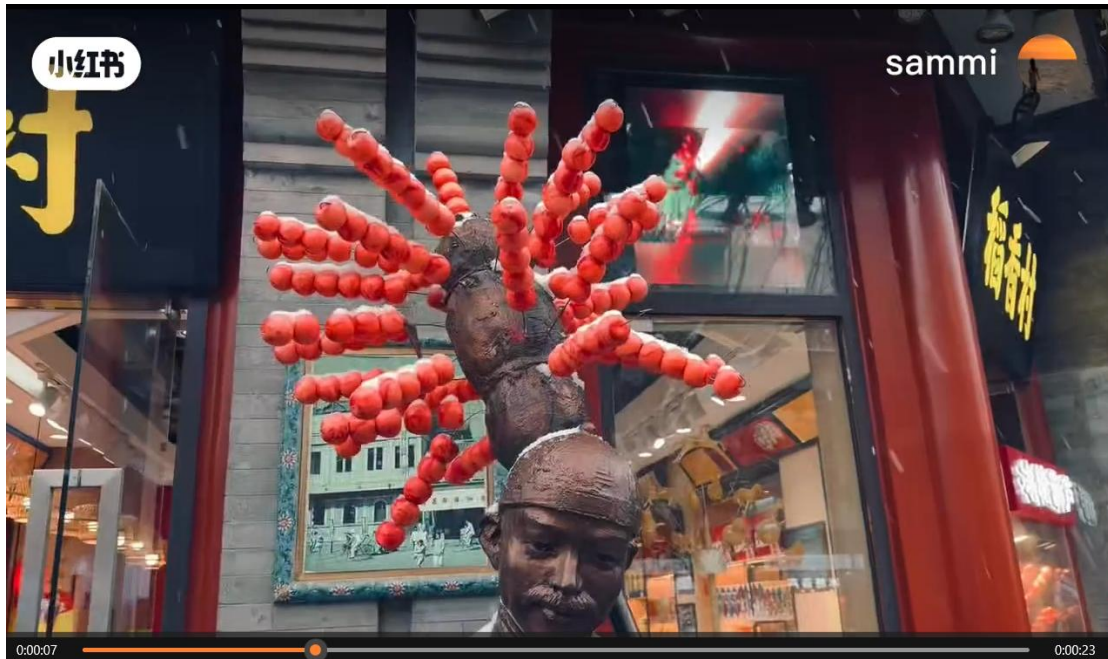


Fig. 13

The short video also skillfully uses sounds of the subway carriage as an emotional cue during a transitional moment in the video (see Fig. 14). When the visuals shift from snow-covered streets to the subway interior, such sounds initiate the emotional variation. The familiar noises of door-opening, subway announcement voices, and the ambient murmurs of passengers forge an enclosed, orderly, and cozy auditory space. This sound shift enhances the textured structure in urban settings and symbolically guides the viewer from the chilly outdoors into a sheltered world of public transportation. In this moment, sound effects signify an emotional transition, and more importantly, unveil the warmth and a sort of sanctuary sense permeating the city, labeling the “winter in Beijing” not merely as a visual presence but also a profound, tangible experience.



Fig. 14

Overall, the video purposefully adds and accentuates indoor voices and subway sounds to transform the winter street scenes into a narrative of emotional warmth. Sound effects transcend beyond tools of reality recreation to become creative forces that shape emotional structures. They lend psychological warmth to urban space, add emotional depth to Beijing's city image, and enable viewers to experience the city's humanistic concerns and human bonds.

In Xi'an's short video "*I Want to Walk with You from Morning Bell to Evening Drum*", the creator threads the ticks of a clock's second hand throughout the video. This subtle yet symbolically rich sound, as the primary auditory element in this video, guides the audience to sense the city's rhythmic pulses and emotional flow through the hearing experience. Unlike the bustling street sounds or vibrant chatters common in real-life urban settings, it is by no means loud or complex. Its simplicity and steady pace endow it with a dual role of signifying time and facilitating emotional continuity. Meanwhile, the visuals record periods in the city's daily life, from the street-wide bell sound echoing the tranquil early morning to the busyness of evening crowds.

Gradual shifts in these scenes present a flowing portrait of Xi'an's daily rhythm. The soft ticks of the second hand remain constant throughout this camera sequence, invisibly threading the changing visuals. This consistent, measured sound seed in viewers the passage of time even when the images shift. Sounds applied in this case, aside from supporting the visuals as their background details, emerge as symbolic spaces that shape the city's temporal experience and perceptibility. In this way, the poetic idea of "morning bell to evening drum" is translated into a tangible auditory experience.

In addition, ticks of the second hand lead to a more internalized rhythmic expression of the city. Unlike traditional musical rhythms that often create emotional tension, the tick represents a gentle, steady pace that signals a sense of "slow yet continuous" temporal flow. This kind of rhythm carries strong psychological implications, subtly indicating continuity and a sense of order in city life. In a historically and culturally textured city like Xi'an, time both indicates the present moment and symbolizes the coexistence of past and present. Ticks of the second hand, as the smallest measurable unit of time, reinforce the fusion of daily life and historical continuity, elevating Xi'an from a mere spatial setting to the so-called "time container" with emotional depth.

By selectively amplifying the second hand's ticking sounds and weaving its rhythms throughout the video, the creator merges time itself into the city's symbolic meaning. This minimalist, precise sound strategy not only enhances viewers' temporal sense and emotional depth of Xi'an's city image but also demonstrates how specific sound effects in short videos can transcend real-world representation by bolstering emotional symbolism and rhythm construction.

In the short video "*This Half Minute is the Taste of Shanghai*", the creator densely and rhythmically combines the curated sound effects to construct an auditory path that spans across decades. The audience can be plunged into three decades of the city's

evolving identity within just thirty seconds. Instead of applying voice-over or explicit narratives, the video employs rhythms, layering, and symbolic functions of sounds to convey historical evolution and cultural emotions. In this context, sound effects are not simply background noise but also the symbolic language to drive the narrative, convey feeling, and evoke memory.

The video opens with a sharp, fast-paced percussive sound, reminiscent of a ticking clock (see Fig. 15). This rhythmic, attention-grabbing effect, instead of being extracted from a real-world environment, is artificially made to function as a “temporal device” signaling the journey start through time telling and the awakening of collective memory. Its usage of sounds alone establishes the emotional tone and immediately places the viewer within a historical context deep-seated in the past century. This creates the first symbolic layer of “time travel,” setting the stage for the unfolding narrative.



Fig. 15

Following the opening sections, the video introduces a sequence of sound effects closely tied to the city’s physical environment, such as the screech of old car tires, the footstep sounds in a crowd, the jingle of bicycle bells, and the iconic chimes of the

Fairmont Peace Hotel. These sounds that are strongly associated with specific areas evoke the viewer's imagination of traditional Shanghai street scenes and serve as auditory symbols of daily rhythms glowing in the past. The video applies fast-paced editing to render layered and speedily switched sounds. They form a complex soundscape that condenses the auditory portrait of the city's history. The city's evolution thus comes across as an "audible" timeline as empowered by sound design.

The creator incorporates natural sounds such as wind and rain into the video, adding further texture to the soundscape, and interweaves them with mechanical and artificial sounds to create a kind of "compressed archive" of urban memory within just half a minute. In this soundscape, the viewer does not simply move through space, but audibly replays the city's history.

This Shanghai-based video case showcases the sound-enabled construction of the city's historical and emotional structure by introducing a fast-paced beat, layered iconic ambient sounds, and extending the mood expression through natural environmental audio. Specific sound effects function as symbolic vessels of time and memory, leading to shaping a dynamic, textured, and rhythmically engaging city image of Shanghai. These auditory features deepen the viewer's emotional immersion in the city's cultural atmosphere and highlight the symbolic power of sound in urban visual storytelling.

Short video case studies of Beijing, Xi'an, and Shanghai show that specific sound effects have risen in prominence amid city image construction, carrying symbolic meaning and shaping the city's perceptual experience. The combined use of rhythms, sound expansiveness, and associative power of sound enables the effective conveyance of abstract emotional tones, cultural atmospheres, and historical depth to the audience. In this way, sound effects become an irreplaceable symbolic medium in short video narratives (Görne,2019).

In terms of symbolic sound strategies in short video production, Beijing's case emphasizes the emotional warmth of cityscapes. Through skillful orchestration of sound effects, the video evokes a sense of gentleness, familiarity, and human connection to the city atmosphere in viewers. Xi'an's case further stresses the use of rhythmic auditory cues to build a temporal sense and psychological pacing, presenting the city as stable, orderly, and continuous. Shanghai's case highlights fluidity and layering in its sound design, constructing a distilled narrative of city memory using sound effects. This imparts to the city image a sense of sound-perceivable historical depth and cultural richness.

The strategy differences of symbolic sound effects in the video cases of these three cities reflect the diverse, strategic functions that sound can undertake in urban visual narratives. Sound can work as the emotional vessel, means of emotional and rhythmic representation, memory arousal tool, and the meaning-making medium. It's no longer confined to being an accessory to the visuals. Rather, it serves as an essential part of the city's emotional structure and symbolic system. Sound design imbues the city images with a more richly layered auditory construction and symbolic expression, allowing for a more diversified and engrossing representation.

Voice-over in city image construction

In short videos, voice-over, as a non-diegetic sound element, is substantially subjective and purposive, and supports creators in engaging with the narrative, expressing attitudes, and guiding viewers' comprehension. The authenticity-enabled background music shapes the atmosphere or ambient sounds, while voice-over is more directly linked to certain language, emotion, or thought. Voice-over, in addition to containing explanatory information, helps with shaping the narrative rhythm and

emotional trajectory. This sound format holds significant potential in personalized expressions of city image construction.

In a fast-paced media environment flooded with short videos featuring rapid cuts and condensed information, voice-over is vital to structuring and organizing conveyed meanings. Creators can turn fragmented clips into coherent narrative units with a sense of rhythmic pattern and logical flow (Matamala, 2018) through choices in sentence sequence, speech pacing, and transitions. This, in turn, helps establish a guided “viewing path” on the auditory level, where voice-over directs viewers’ attention, interprets visual content, and offers emotional cues, creating a coherent flow that guides them through fragmented scenes.

Voice-over is often displayed through the first-person narrative, in forms of personal reflection or poetic language. This injects emotion and memory into the urban space, transforming the city from a merely observable location into a sensorily perceivable emotional landscape. At the acoustic level, it shifts the functions of sound from visual interpretation to emotional guidance, and from objective representation to subjective experience. As a result, city images in short videos embody greater warmth, rhythm, and individuality (Shao et al., 2018). In this way, voice-over helps turn city representation into a more immersive and emotionally evocative experience.

Emotional expression through voice-over

This section focuses on the roles of voice-over in city-specific emotion expressions, by exploring how tone of voice, narrative rhythm, and personal feeling are used to make Chinese cities both “visible” and “perceptible”. Voice-over, when blending the above techniques, auditorily intensifies the emotional impact of the city image and deepens viewers’ city perception, proposing a more nuanced and affective experience of the urban environment.

In the short video “I Bet You’ve Never Seen Beijing’s Friday Rush Hour”, the voice-over forsakes the traditional narrative but emerges from a real-time conversation between two people inside a car. Though such spontaneous, context-based dialogue lacks literary or rhetorical polish, its authenticity and raw subjectivity construct compelling emotional impacts. Viewers, when appreciating the tone of voice, speech rhythm, and dialogue details, don’t just learn about the city’s traffic conditions through hearing, but also genuinely experience the emotional journey in a high-stress environment.

The video was entirely filmed from the fixed viewpoint of the car’s interior, enabling viewers to watch the traffic stream through the windshield. The visuals remain minimal and restrained, but it is the conversation sounds that bring tension and emotion to the scene (see Fig. 16). The driver and passenger discuss the road conditions, lane choices, and the stress of traffic navigation. Shifts in speech speed, rising intonation, repeated confirmation, and exclamation all contribute to a sense of being “caught in the rhythm of the city”. As they chat about changing directions or navigating the challenges of rush hour traffic, tension, frustration, and resignation delivered through their voices make Beijing’s peak hours emotionally relatable.



Fig. 16

In this case, the voice-over conveys abundant information and goes further to actively shape the emotional atmosphere of the city. Unlike ambient sounds or background music that typically set moods, this highly subjective representation of “natural language” reflects an individual’s psychological state more effectively within a specific urban context. Such voice-overs vividly communicate part of the city’s character, which encompasses the coexistence of order and disorder, and the overlap of efficiency and anxiety.

The utilization of conversational, everyday dialogue by the creator adds a layer of personal feelings on top of the visual imagery, imbuing Beijing’s city image with a clear emotional tone that extends the visible scenes. The subjective voice-over and its rhythmic delivery become key to constructing the city’s psychological space. As a result, the city is not just a landscape composed of physical elements, but also a social environment filled with emotional resonance and lived-in experience.

In Xi’an’s short video, “*Tonight, the Lights Shine for Miles—No City Is More Striking Than Chang’an*”, its voice-over features the first-person narrative running throughout

the video. It functions as both the structural backbone of the narrative and a key tool for conveying the city's emotional atmosphere. This narrative is marked by strong subjectivity and emotional intensity, in contrast to those portrayed by objective, explanatory voice-overs. The voice-over, through its speaker's rhythm, tone, and emotional progression, creates an auditory journey moving from spatial exploration to emotional immersion, gradually engrossing the viewer in the city atmosphere through language design.

At the video's start, the narrator depicts the experience of ascending the ancient city gate in an unruffled, reverent tone, simple language, and measured pace. At this stage, the voice-over lends primary support to the visuals, providing spatial orientation and guiding the viewer's gaze into the video. However, following shot shifts to scenes of the vibrant night market and street food, the voice-over begins to accelerate its rhythmic flow and pick up a more expressive tone (see Fig. 17). Excitement and enthusiasm that emerge from the narrator's voice-over reflect a genuine emotional response to the city's animated street life. This calmness-vigor emotional progression forcefully conveys the warmth and vitality of the city to the audience.



Fig. 17

In the latter part of the video, as night falls and lights illuminate alongside the ancient architecture across the city, the narrator shifts the voice-over's emotional tone, reflecting on the historical grandeur of ancient Xi'an when illustrating the nightscape. Subtle pauses, slight voice tremors, and a gradual rise in emotional intensity reveal a deeper sense of connection through the voice-over. This narrative style, where emotions emerge naturally from the scene, blends the present-day urban experience with the imagined memory of its past.

As a result, the voice-over moves beyond interpreting visual impressions to shaping an emotional resonance bridge between past and present. Xi'an is crafted as a confluence of personal emotion and cultural memory. The sound design reinforces the city's historical richness and instills symbolic weight and emotional depth into it.

In the short video "The Dreamer Jiang Shiyi: What Kind of City Do You Think Shanghai is?", the first-person voice-over creates a highly personal and intimate soundscape. Such a means of narrative, in contrast to objective or poetic voice-overs, blends individual experience, subjective emotions, and city memories. Shanghai gradually comes across as warm, textured, and emotionally fluid due to this sound language. The voice-over, apart from guiding the viewing experience, encourages the audience to sense the city and fosters in them the emotional resonance grounded in everyday life and lived experience.

The video opens with a straightforward question: "Do you long for Shanghai?". This opening immediately breaks away from a conventional descriptive tone and cultivates a sense of proximity between the speaker and the viewer, driving the audience to introspect on their perceptions and thinking of Shanghai. However, this question, which remains unanswered in the video, serves as an emotional trigger that draws the viewer into the urban space and personal memories familiar to the creator.

The creator brings in her own memories and emotional experiences, tenderly narrating moments of uncertainty and self-struggle as she interacts with the rhythms of the city. Despite emotional weight born by these reflections, she maintains her calm tone and steady voice (see Fig. 18). This balanced narration of emotion within restraint enhances the expressive power of voice-over as the sound language, and allows the city image to rise above surface-level aesthetic, depicting city as an inner space where rich emotions can be articulated and navigated.

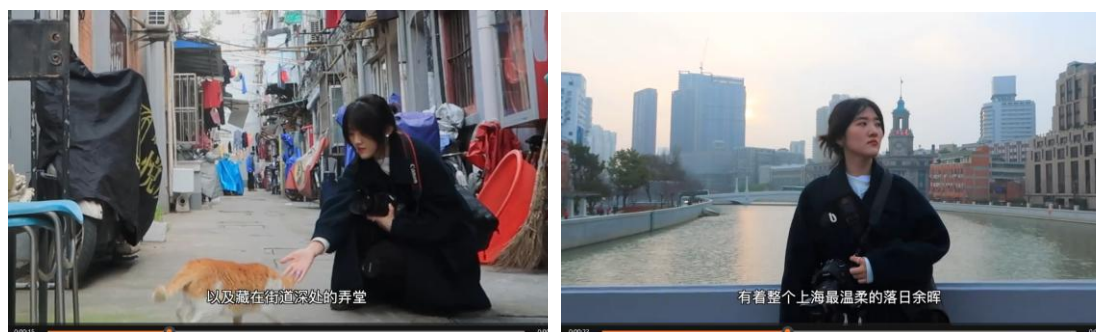


Fig. 18

The voice-over shifts from raising a question to making responses, and from the real world to envisioned possibilities of life near the video's end. Although the creator candidly shares her mental fatigue and sense of uncertainty, she gradually shifts her tone by injecting serene strength and hope into the voice-over. This transition from emotional struggle to a more positive outlook redefines the city not into a passive backdrop or memory container, but a space capable of receiving, transforming, and inspiring individual emotions. At this moment, the voice-over completes its shift from observing the city from the narrator's first-person perspective to emotionally connecting with it, turning the narrative into a shared experience between the speaker and urban spaces.

As inferred from this short video, the creator applies a subjective, sincere, and emotionally layered voice-over to shape Shanghai into a city rich with psychological connotations.

Firstly, beyond conveying personal experience, the voice-over navigates the viewing experience through its emotional shifts, tone variation, and open-ended language to empathize with and interpret the city's emotional tonality. Secondly, this context also features the voice-over as both a storytelling device and the warmest and most emotionally resonant element within the city's emotional structure. Thirdly, the voice-over expresses emotions and works as an auditory mechanism that transforms the city from a spatial vessel into an object of psychological projection. Consequently, it marks a narrative shift from "observing the city" to "being emotionally touched by the city."

According to the short-video case studies of Beijing, Xi'an, and Shanghai, voice-over, as a key approach to emotional expression, demonstrates strong subjectivity and symbolic depth in image construction. In spite of the varying language styles, narrative rhythms, and emotional intensity, the videos all represent multiple functions of voice-over, including emotional initiation, psychological transition, and affective connection.

In Beijing's video case, the voice-over is expressed through casual, real-life conversations, and its immediate and natural emotional expression highlights genuine reactions to stress in urban life. This narrative style emphasizes the presence of lived-in experience and imparts the tangible affective warmth to the city image.

In Xi'an's video, the voice-over follows emotional progression, with shifts in rhythm and tone that reflect the evolving relationship between the creator and the city. Emotions seep through the voices progressively, constructing a sense of historical

depth and imaginative space. In terms of Shanghai's video case, it takes voice-over as a means of emotional reflection and identity expression. Its composed rhythm and gentle tone reshape the city into a psychological space that holds memories and hope through personal experience.

Voice-over design in the three cities' short videos reflects sound strategies distinct in nature: voice-over presented by immediate responses in Beijing, by voice-over emotional progression in Xi'an, and by introspective expression in Shanghai. Despite their format differences, all three cases demonstrate the vital parts of voice-over in transforming objective urban space into a subjective emotional landscape. The addition of human speech remodels each city as a mere visual backdrop into a place to be narrated, felt, and emotionally interpreted. In this sense, voice-over design shifts the representation of the city from a spatial display toward an emotional identification, deepening the connection between audiences and the urban image.

Voice-over and narrative progression of short videos

This section exemplifies how creators apply voice-over strategies to drive the development of content in short videos portraying Beijing, Xi'an, and Shanghai. It is devoted to how voice-over organizes language, manages pacing, and builds logical connections across scenes. These sound-based approaches help viewers better navigate and comprehend the urban spaces, cultural characteristics, or personal experiences presented in the videos. These techniques also ease viewer's way with following the creator's narrative flow and engrossing themselves in the unfurling storyline.

In the short video "*In Winter Beijing, I Heard the Ancient Roar of Loong (Chinese Dragon)*", the voice-over plays a central role in shaping the overall narrative structure. Revolving around the theme of "Loong search", the video visuals depict the

creator walking through the city with scenes captured in real-time, while the voice-overs advance the real-life storyline. Voice-over, by rendering its semantic guidance and rhythmic control, helps the audience understand this cultural exploration as a journey to uncover the Loong imagery embedded in Beijing. It is this narrative layer that facilitates the coherent, complete city narrative through the short-format video.

The creator begins the video by introducing a reflective and steering voice-over, posing questions, and stirring up the viewer's curiosity. This initial narrative quickly establishes the core content focus and charts a clear course for the sequence of ensuing visuals.

During shot transitions to different locations, the voice-over continues to thread the entire story, capturing symbols and urban spaces associated with Loong imagery. In one aspect, it supplements the visuals with cultural contexts unable to be merely conveyed by images, and elucidates the historical origin, symbolic meanings, and spiritual significance of Loong. In another aspect, it serves as a narrative bridge across temporal and spatial transitions, helping maintain the coherence in videos comprising visually fragmented scenes.

Despite the frequent on-screen presence of the creator, the fast-paced nature of short videos determines the brief duration of each shot, suggesting the difficulty in delivering a continuous narrative through live speech or on-camera actions. In this context, voice-over, as a stable and controllable storytelling device, is fundamental to retaining the coherent logical flow driving the narrative content forward, and shifting it from concrete scenes to abstract concepts.

This process sees how the “Loong search” evolves from a spatial journey into a symbol of cultural exploration and emotional connection. In Beijing's short video case, voice-over functions not merely as a supplement to meaning but also as the

driver for the story development. It sustains its function as the main thread of content, connects fragmented scenes at the structural level, and regulates pacing to smoothly navigate viewers all across the video. This sound approach allows for the presentation of a clear theme, a coherent structure, and effective content delivery in a short-format video. It highlights the potential of voice-over to transform from delivering simple expressions to taking on a narrative function, demonstrating the irreplaceability of sound language in shaping city imagery.

In the Xi'an short video "*Shooting Warring States Robes in Xi'an*", the voice-over mainly consists of the photographer's directive language. This type of speech, instead of operating as a traditional narrative or personal reflection, features a series of highly practical instructions such as "Please turn around." "Now please look up." or "Please hold it." that guide the model's movements and poses during photoshooting (see Fig. 19). From the structural dimension, despite being straightforward in nature, such language plays a distinct role in advancing the content, turning what could have been a sequence of static visuals into a continuous process of action.



Fig. 19

The video, centered around an artistic photoshoot process in Xi'an, depicts the model who shifts her posing in various settings through changes in styling and position. Due to the absence of a linear storyline or event progression, it's easy for viewers to consider the visuals alone fragmented or insufficient in rhythm. The photographer's spoken commands help bridge these gaps by creating an inter-movement logical flow. Each instruction, paired with a subtle shift in the frame, develops a sense of continuity. In this way, the voice-over becomes an "operational cue" that connects individual shots, allowing the viewer to follow both the filming progression and the subtle narrative being built through visual transitions.

Additionally, this voice-over form, empowered by real-time interactions, invests the short video with a sense of "on-site documentation." The content shown in the short video presents both the outcome and the creative process. The presence of voice-over uncovers the deep-seated logic of action that may otherwise be compressed in the visuals. As a result, the viewing experience isn't confined to static image display but emerges as a language-guided dynamic structure marked by physical movement. The audience, rather than being passive observers admiring the photographed images, gains insight into the creative process through the photographer's verbal cues, enabling a more in-depth understanding of the content.

Overall, voice-overs in this short video center on directive language. Rather than delivering information through explicit narration, it pushes the visual sequence forward by facilitating the subject's actions. This sound strategy helps shape a clear viewing path, enhances the video's rhythm and coherence, and demonstrates the key role of voice-over in unfolding information within non-narrative content.

In the short video "*In Shanghai, Searching for the Memories of Silent Separation Seven Years Ago*", voice-over permeates the entire narrative, functioning as a key mediating instrument that bridges real urban spaces and fictional memories derived

from film and television. Unlike conventional spatial display or visual montage, the video constructs a lucid path of situational transitions with voice-overs. As viewers follow the creator's steps into diverse locations, they undergo a narrative journey shifting from reality to story, and then circle back to the real world. The voice-over provides more than just contextual framing, since it effectively guides the pacing and rationale of the viewer's experience, actively steering the rhythmic and logical progression of the viewing experience through its semantic structure and tonal modulations..

The video commences with the line, "In Shanghai, I remember the memories of Silent Separation", which immediately anchors the theme and serves as a narrative portal. This concise line of expression dualistically operates as an emotional guide and the structural entry point, leading the viewer from a real-world state into a memory-framed revisiting process enabled by a television drama. The visuals then traverse several locations in Shanghai, spanning M50 Creative Park, Antique Cafe, and later the venue of a celebrity-based cultural festival.

Despite the physical separation of these places, the voice-over consolidates them into a continuous narrative trajectory. Each spatial shift is introduced with concise, purposeful lines for clear navigation. In this way, the voice-over emerges as the enabling element that connects the locations and sustains the video coherence from start to end.

Notably, the voice-over assumes a "withdrawal" function in the final part. The creator, forsaking specific spatial descriptions, shifts to a brief personal reflection, concluding the video with introspective language. This transition to inward introspection marks the viewers' return to the "city in reality" from the "city within the story". Such voice-over usage thus concludes a structural loop where it facilitates

not only a spatial transformative journey but also a full cycle of situational transformation for viewers.

In Shanghai's video case, the tripartite structure in the voice-over, namely introduction, development, and withdrawal, effectively connects space, content, and narrative rhythm within the short video's temporal constraints. The voice-over, beyond taking on the cueing function, uses purposeful linguistic organization and controlled cadence to position itself as a key driving force for the overall video progression. This approach transcends the role of voice-over as an emotional vehicle, pushing it towards being a structural device that contributes directly to the construction of meanings, reinforcing its value in shaping city representation within short video storytelling.

The above analysis of the short video cases from these three cities demonstrates that although the voice-overs differ in form and language style, they all play a structural role in guiding content development in videos. In each case, the voice-over fills the gaps unable to be fully covered by the visual elements alone. It establishes a clear narrative flow, enhancing the storytelling coherence and completeness.

In Beijing's video case, the voice-over uses structured language to connect spatial shifts with the unfolding of conceptual ideas, allowing the video theme to smoothly advance across different shots. In Xi'an's video, directive language reveals the behind-the-scenes photoshoot, helping viewers grasp the continuity of actions and the underlying transition rationale. Shanghai's video case is contingent upon compelling, guiding, and transitional language to shift the video setting from real-world urban locations into fictional storytelling contexts and next circling back to the real world. The voice-over functions as a content cue, navigating the viewer's journey throughout the video.

In addition, voice-overs in the short videos studied in this research embody different functional focuses in the progression of video content. The layered organization of thematic language is stressed in Beijing's case. Xi'an's video case highlights the sequential presentation of process-related information. As for Shanghai's video case, it focuses on delivering structural closure through proficient control of situational transitions. These strategies together demonstrate that voice-over, as an accessory sound element, indeed plays structural roles in short videos. It is helpful in advancing the video content, organizing the viewer's interpretive flow, and coordinating rhythms for visual transitions. Consequently, voice-over has become an essential vocal mechanism in the construction of city images.

Conclusion

A systematic analysis of multi-faceted sound expression in short videos depicting Beijing, Xi'an, and Shanghai indicates that creators share certain commonalities in their creative use of sound strategies, but also demonstrate distinctive approaches inspired by the unique attributes of each city.

Overall, creators from all three cities place pronounced emphasis on the roles of sound in guiding emotional responses, regulating narrative rhythm, and integrating narrative structures. The combined application of background music, ambient sound effects, and voice-over constitutes a key sensory pathway in constructing city images within short videos. In a fast-paced yet fragmented media context, sound is crucial to bridging visual jumps, directing the viewing pace, and evoking the affective responses connected to the city. This sound-driven approach enhances the emotional impact of the city's representation and suggests new possibilities for more resonant, engaging communication.

Creators demonstrate sound strategies that align closely with the character and atmosphere of each city. In Beijing's video cases, sound design follows a comparatively more restrained, orderly pattern, emphasizing a historical sense and air of authority, as well as the solemn background music and voice-over featuring clear, rational language and steady rhythm. In the short videos portraying Xi'an, the creators are more inclined towards blending sound effects with traditional melodies, thus conveying the city's temporal depth and cultural richness through rhythmic changes. In Shanghai's video cases, sound expression adopts a more urban and personal tone and integrates faster pacing and lighter intonation. This results in a soundscape that mirrors a modern city image composed of fluid, diverse, and emotionally nuanced identities.

These similarities and differences not only reflect the creators' aesthetic judgments in applying sound strategies but also reveal their insights into and approaches to unveiling the cultural characteristics of each city. A city's image isn't simply shaped by a direct visual presentation of space, but also by purposeful sound organization, whose emotional, rhythmic, and structured qualities are formulated during image construction. The creative decisions underlying sound design also point to deeper variations in how cities interact with digital communication, reflecting distinct mindsets of cultural expressions and communicative contexts within the digital media landscape.

This chapter, combined with the earlier analyses of representational techniques such as color or light and shadow, perceptually prepares the ground of city image construction. Last but not least, it provides essential support for the following chapters that will examine mechanisms of construction from a narrative perspective.

Chapter 6 The Role of Characters in Constructing City Images

Introduction

In short video content, characters emerge as visual subjects key to constructing and perceiving city images. Compared with color, light, sound, and earlier discussed representational techniques that mainly operate on a sensory dimension, character design imparts deeper humanistic concerns to the city's image representation.

Character design and behavioral arrangement help characters build the bridge between urban spaces and everyday life, and between local values and collective memory (Tseng, 2013). Their embodied emotional expressions and social relationships help viewers forge emotional bonds and a sense of affinity with the city displayed on screen, thereby adding warmth and intimacy to the city image.

This character-focused chapter is mainly structured into three dimensions. First, it explores how characters contribute to weaving dispersed urban scenes together and prompting deeper dives into spatial exploration. Second, it examines their representation of city life by portraying daily experiences and local spirit. Third, it delves into ways through which characters evoke reflective perspectives, instructing viewers to reimagine the city itself and individuals' positions within its development landscape. A comparative analysis in this chapter manifests the roles of character strategies in depicting and reinforcing distinctive images of Beijing, Xi'an, and Shanghai, respectively.

Driving the narrative forward

The short-format video context, compared to static shots or simply scenic depictions, features a more coherent and dynamic representation of characters. Short video characters are often tasked with directing the narrative rhythm and organizing the

visual structure (Jing, 2024). Characters' movements seamlessly connect places within the city, designing a distinct viewing path and spatial experience for the audience to visually follow.

As visual anchors for shot transitions whether between neighborhoods or indoor and outdoor scenes, characters bridge these locations and construct a more fluid viewing experience (Magliano et al., 2011). Their presence furnishes narrative contexts to various parts of the city, breathing life into the otherwise isolated spaces and adding rhythm and emotional warmth.

Beyond linking fragmented parts of the city, characters also help navigate in-depth explorations depicted in short videos. Their footsteps direct viewers into city corners that are deeper-rooted in everyday life, and local cultures that extend beyond iconic landmarks or areas familiar to tourists. This “immersive” mode of guidance infuses city image construction with a more subjective tone and affective depth.

Threading city scenes

It's necessary to alternate shots frequently in the short-format video, in that a city's spatial structure and overall ambience cannot be covered by a single scene. In this context, characters are essential to threading various shots. Creators, by applying the movement paths, behavioral logic, and gaze direction of characters, manage to visually connect otherwise scattered locations, rendering a more cohesive, continuous city image for the viewers. This section analyzes how creators leverage character design to thread urban scenes in these three cities, thereby maintaining coherence in the constructed city image.

In Beijing's short video “Traditional Beijing's New Year is Amazing”, the creator, who also acts as the character, captures the city's festive ambience during the Chinese

New Year by displaying a range of diverse scenes. She effectively links different locations across the city through her on-screen appearance, portraying a total of eight representative festive scenes like museums, industrial parks, and hutongs to epitomize Beijing's spatial diversity (see Fig. 1).



Fig. 1

Regardless of rapid transitions of multiple festival scenes and locations, or changes in the background, the character remains at the frame center in a front-facing manner, whose presence anchors the storytelling throughout the video as a consistent visual element, offering a clear, stable focal point. This character strategy presents the city image as an organically integrated whole, which reduces the sense of fragmentation resulting from frequent scene shifts and keeps the narrative coherent all across the viewing experience. This character-enabled visual consistency operates as a key narrative thread across diverse settings, enabling a coherent representation of the city image characterized by a natural flow of festive vibes across spaces in Beijing during Spring Festival.

The creator adopts a visitor's perspective in character design, drawing viewers into a perception framework centered on emotional engagement through her personal experiences. As the character senses the festal ambience as an outsider in Beijing, she leads the audience to focus on her subjective emotional responses. This approach brings an emotional identity to the city representation, easing viewers' way in relating to the character's emotional journey and thereby acquiring a more intuitive

understanding of the Spring Festival celebrations in Beijing.

The blend of visual consistency and individual experience expression helps the video establish a sensory rationale in which the character is engrossed in the city, and the festive ambience seems within easy reach. This short video character, who visually and consistently threads varied scenes, shapes organic connections between city spaces. Meanwhile, celebrative moods conveyed from the character's lens lend a marked warmth and cultural spirit to the city image. In a word, this effective character strategy balances visual coherence with affective expression in constructing Beijing's city image.

In Xi'an's short video case "A Tribute to This Encounter", the creator develops the narrative by touring along the city's Metro Line 2 route, filming a series of still frames in front of station entrances. This video successfully constructs a sense of citywide continuity by recording shots of the character striking poses with each station's name sign. The selected locations, such as Tiyuchang (Shaanxi Stadium) Station, Anyuanmen Station, and Daminggongxi (Daming Palace West) Station (see Fig. 2), highlight Xi'an's historical axis and illustrate a spatial logic extending from core tourist attractions to daily inhabited zones. Despite their geographical distances, the characters' consistent on-screen appearance substantially diminishes the sense of spatial separation and creates a visually coherent urban rail route.



Fig. 2

The short video adopts a highly simplified yet distinctive compositional strategy. In

detail, the character's consistent stance at the frame center pushes the visual focus towards the metro station signage. This strategy minimizes environmental background details but pairs the character with signage to jointly emblemize the city's recognizability. In addition, the character is assigned a different pose at each station, introducing visual variety and playfulness on the basis of compositional structural consistency. This continuous fusion of people and landmark signs functions by visually anchoring the video narrative, which naturally connects the city's spatial layout with the character's repeated appearances.

The character plays to activate urban scenes and organize the visual structure. The character, standing in front of the non-sentient metro signage system, brings life to space through her presence in the shots, transforming metro stations from simple transportation nodes into meaningful representations of the image of Xi'an. This creative strategy associates public urban spaces with individual perception, blurring the boundary between city locations and individual experience. The character's visual presence and physical expression facilitate an interactive city image construction positioned between spatial organization and cultural sensibility.

In the short video "I Came to Shanghai to Find the Filming Locations of *Light My Fire, Warm You*", the creator applies the idea of "revisiting TV drama locations" as a narrative thread, linking multiple culturally significant landmarks in Shanghai through the characters' repeated appearances. This short ingeniously draws on materials from the television series *Light My Fire, Warm You*, whose characters' actions imbue locations such as Jiuli Workshop and the banks of Suzhou Creek with emotional layers.

The video creator himself functions as another character who enters these places as a real-life viewer. Such a pairing enables otherwise scattered city spaces to be consistent in their visual structures. Within this creative framework, the character

serves as a spatial connector and a visual intermediary that translates media memory into lived perception, unleashing and sorting out the meaning of each scene (Oatley, 1995).

Beyond operating short video characters as fixed visual elements, the creator builds a relationship of “inheritance” and “reproduction” between characters and the city space. Aforementioned TV series characters endow Jiuli Workshop, Suzhou Creek, and places alike with symbolic narrative meanings (see Fig. 3) through their actions. The presence of the short video creator himself reawakens and reinterprets these figurative associations. Spaces once shaped by fictional narratives are re-experienced and reactivated in reality through the masterful use of juxtaposition, imitation, and responses between character and location. This demonstrates how city spaces continuously acquire new narratives and viewing pathways through the persistence of mediated memory.



Fig. 3

In the same video, rather than foregrounding the identities of the TV drama’s characters, the creator focuses on conveying emotional tones expressed by them. The creator often mirrors the positioning of original characters in composition design, making the character an important intermediary between fictional narrative and real-world space. This study, built on the symbolic function of characters in city representation as noted by Liu et al. (2024), further argues that such positioning strategies of characters in short videos directly contribute to the reinvigoration of

cultural memory through embodied spatial performance.

These three video cases demonstrate how short video characters connect city spaces and construct city images as strategic instruments from diverse perspectives. Despite variations in cases concerning the creative theme, scene selection, and character design, the three short videos all highlight that characters perform the role of a structural thread running throughout the visual narrative. This underscores the universal value of character-based strategies when constructing city images via short video content. As for the visual representation of specific cities, character strategies embody a high degree of flexibility and adaptability, as unveiled by different ways of presenting characters, organizing scenes, and approaching cultural perception by creators in each city.

However, while the above creators all connect spatial layout with characters, each case distinguishes its emphasis on construction rationale. Beijing's video case focuses on the character's visual stability and role in guiding viewers to sense emotional tones, emphasizing the recreation of everyday scenes. Xi'an's video case stresses the juxtaposition of characters with symbolic elements of the city's spatial structure, presenting a well-defined sense of spatial order.

In comparison, Shanghai's video case applies characters to engage with mediated memory, navigating the viewing experience of perceiving the city's cultural richness through a continuous interplay between fiction and reality. These strategic differences dualistically unfold the multiple functions that characters serve in visual organization, and the diverse pathways of city image construction shaped by varying cultural and physical contexts.

Delving into city exploration

Creators use characters as viewers' guides in perceiving profound city textures, other than filming panoramic overviews and speedy glances. This exploratory mode of presentation moves beyond signature landmarks and popular attractions by permeating the inner fabric within the city. Characters' movements and roles are turned into key entry points for viewers to observe and engage with the city image (Xi, 2024) being constructed in the video. This section analyzes how the guiding role of characters is applied to reveal the city's in-depth spatial structure and cultural context, thereby enriching the complexity and depth of city images shaped by short videos.

The renowned modern Chinese writer Lao She is directly referenced by the creator of the short video "This Moment in Beijing is Heaven on Earth Once Praised by Lao She" in its title. The video draws upon the delicate portrayal of everyday details in Lao She's literary works, which serves as a distinctive narrative perspective deeply anchored in Beijing's cultural heritage and leads viewers into more profound explorations of the city in autumn. This strategy translates the city image in his writings into the contemporary visual language, elevating the character into both a cultural reference and a reconstructed mode of observation, and shaping a narrative pathway in which the city is revisited through the lens of a cultural figure.

The creator recreates Lao She's observational perspective in Beijing's autumn through a vivid visual composition. The shots of hutong alleys, along with people and small-sized animals strolling under the sunlight, reflect a precise attention to detail (see Fig. 4). By introducing the character's perspective, the video visuals offer a reflective, culturally informative viewing journey that breaks free from straightforward documentation, hence presenting a city image impressively marked by aesthetic order and cultural diversity.

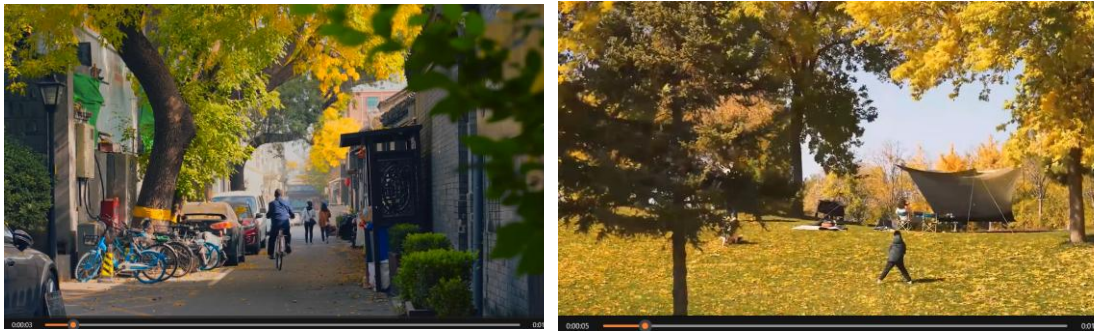


Fig. 4

This character strategy enables a shift from literary text into visual representation in the constructed city image. The sensory details and linguistic texture found in Lao She's works are distilled into compelling visual language. A well-known line in one of Lao She's masterpieces, *A Dream of Dwelling*, reads, "The sky stretches blue, high above, clear and silent." Although the video doesn't mention the line explicitly, its visuals meticulously recreate the serene atmosphere it evokes in viewers with orchestrated sky tone, upward shooting angles, and muted use of ambient sounds. This text-to-image transition stands as a visual re-creation of the literary perception, rather than a superficial imitation. In this way, viewers are engaged to resonate unconsciously with the city as once envisioned by the writer Lao She.

The creator unveils a deeper discovery of Beijing's autumn imagery and stirs a resonance between urban space and literary memory by introducing Lao She as a famed cultural figure into the video. In this context, the character's perspective is translated into a cultural guide to enable viewers to reinterpret the city image when immersing themselves in the visuals. This character construction strategy demonstrates a distinctive pattern of cultural depth, turning the city from a passive observed entity into a textured space rife with literary atmosphere. It also functions as a key pathway for city image construction to undertake deeper cultural meaning and personal experience that move beyond shallow visuals.

In the short video "This is Xi'an, also Chang'an", the creator takes the juxtaposition

of “modern Xi’an” and “historical Chang’an” as the central theme, using the visual transformation of characters to direct viewers toward gaining insights into the city’s dual identity. The video places characters at the narrative center, interweaving contemporary life with historical culture to present a multilayered city image simultaneously mixed with multiple temporalities (see Fig. 5). In this context, characters visually guide the temporal periods in history.



Fig. 5

Characters appear as ordinary citizens in the contemporary-styled section of this video. They are dressed in everyday clothing and reflect the genuine lifestyles of present-day Xi’an, appearing in street dining or cycling along ancient city walls. These characters represent Xi’an’s advanced modernization and sketch a portrait of its contemporary look. In contrast, these scenes are then followed by characters wearing traditional attire in the ancient-style section, where an actor plays the role of the renowned historical poet Li Bai, and tourists all dress in historical costumes.

This evokes in viewers the temporal sense of transition and prompts viewers to enliven the cultural memory of “Chang’an”. This visual shift in character design concerns not only stylistic change but also a culturally charged vessel of visual guidance that allows the audience to fluidly alternate between the two urban imaginaries and comprehend the continuity of the city’s cultural identity during the viewing experience.

The creator assigns different historical dimensions to the same geographical setting by ingeniously shifting costumes, character actions, and narrative contexts, hence transforming urban space into a culturally perceptible temporal field. This strategy enables characters to bear an essential narrative function beyond working as moving elements in the visuals' background. Characters' visual presence and shifting identities expand the city from an everyday space into a confluence of cultural memory and historical imagination. As a result, viewers gradually gain a strong impression and thorough understanding of the city as a spatial fusion of history and modernity.

In the short video “A One-Day Tour in Shanghai Summarized After 30,000 Steps”, the character functions as a “first-hand experienter” who is key to steering the viewer towards delving deeply into the city. The creator appears in the video as the main character, visiting multiple locations across the city in person. The video unfolds a replicable and immersive urban journey through continuous walking and scene documentation under the theme backdrop of “taking 30,000 steps”. While typical checklist-style travel videos merely check off landmark scenic spots with cursory looks, this video case constructs a city navigation route that blends both classic highlights and personalized discoveries that are all empowered by the character's physical movement.

The creator begins by walking towards well-known iconic buildings such as the Wukang Mansion and purposefully adds everyday places along the way to the building, such as cafes, dessert shops, local restaurants, and film locations. These intermediate stops, despite not serving as city landmarks, remain essential components of the city's image (see Fig. 6). Shanghai is vibrantly transformed from an aggregation of destinations into a string of continuous, in-depth experiences through the character's footsteps.



Fig. 6

This character-guided exploration reflects the creator’s well-planned route design strategy. The chosen path, primarily reached through walking and public transport, lowers the barrier to city exploration but heightens viewers’ immersion and accessibility. This character-based strategy helps the creator actively construct a structured framework for discovering the city, where the audience is guided to follow characters’ footsteps and experience various dimensions of the city within a short viewing duration (Du et al., 2022). In this case, a structure, personalized city experience is rendered thanks to the thoughtful route design by the creator, encouraging more enthusiastic engagement from viewers.

As powered by this character-led, path-driven narrative approach, the city image unfurls progressively from isolated spatial and temporal points to a broader landscape. Shanghai, a captivating blend of tradition and modernity, is portrayed through both iconic architecture and scenes of everyday vitality. This content arrangement, which balances structural clarity with lived experience, delivers an organic construction of the city image following the character’s physical movement. It also highlights the guiding role that characters play during urban exploration within short videos.

In the three city cases, the creators all deploy the character strategies to lead the audience into further urban explorations, while their strategic design and presentation styles differ significantly as per each city's distinct features. In Beijing's case, the cultural master Lao She operates as the central guide, directing the audience towards the cultural atmosphere of Beijing in autumn through his literary lens, thereby constructing an immersive path grounded in historical and cultural depth.

Xi'an's case applies both contemporary and historical figures to activate a dual temporal immersion in the city, enabling a smooth alternation from modern everyday life to the historical lifestyles facilitated by traditional culture, reinforcing viewers' perception of this historic city converged by ancient roots and modern dynamism. Shanghai's case presents the character as an on-the-ground urban observer. Through a scrupulously planned route and purposive pacing of scenes, the video organizes a structured and detailed path that inquires into city lifestyles, reflecting the rhythmic charm and openness of this modern metropolis.

Despite variations in the character strategies, a central task is shared by the three cases: steering the audience from superficially perceiving spaces of each city level toward further comprehending its cultural and emotional layers. Characters bear functions of visual storytelling and convey cultural connotations, whether by evoking cultural memory through a literary perspective, enabling temporal shifts through visual transformation, or embedding lived experience into the route design (Liu et al., 2025).

As seen from this dynamic shift from spatial presentation to cultural interpretation, city image construction doesn't just involve the representation of landmarks or scenery but also activates and reshapes the audience's perception of the city's identity and cultural depth through visual means. Each appearance of characters in these

videos can be seen as a renewed engagement with the city's cultural characteristics. In this sense, all video cases ultimately share the purpose of presenting a multi-dimensional city image, though character strategies differ.

Representing city life

In short videos, viewers discover the genuine facets of urban life through characters' speech, routines, and behavior, in lieu of the city's buildings or streets themselves. It is their crucial presence that sublimes city images from shallow visual displays into a more perceptible and relatable experience.

Countless ordinary moments come to constitute the everyday details of city life. Short video creators precisely capture the characters' daily states to lead viewers to cast glimpses into authentic facets of urban living through careful arrangement of daily scenes showcased in the videos, rather than through random recording. As observed by Zhan (2024), micro-film directors tend to approach scene construction with a strong sense of aesthetic regulation and narrative intent. While Zhan's research is geared towards professional filmmakers, this study probes into the endeavors of short video creators who similarly apply visual strategies on purpose to structure everyday scenes, hence directing the viewer's interpretive pathway during city image construction.

Characters, used by creators to implicitly unveil a city's cultural core of the city, help translate abstract cultural qualities into a perceptible visual language through their behaviors, emotional rhythms, and interplay with the urban environment. As argued by Coplan (2008), characters can build an emotional tie with viewers, easing the viewers' way of identifying and resonating with the content. This study further finds that, beyond holding this emotional function, the presence of characters also anchors the city's spirit, transforming abstract ideas into tangible representations.

Depicting everyday city life

Characters express the daily routines in tangible terms, whose day-to-day living situations are exhibited by creators to construct an authentic picture of the city's daily functioning. This section focuses on character-enabled portrayal of everyday city life, and examines approaches to integrating these nuanced moments into the city image construction of Beijing, Xi'an, and Shanghai.

The short video "Listen to the Sounds of the City" forsakes driving the narrative through a main character. Instead, its invigorating portrait of Beijing's daily life is characterized by a group of ordinary people composed of elderly men watching chess play in the park, tourists strolling down the streets, or families resting along the river bank (see Fig.7). Their common motions in life propel the viewing experience where the city's image is no longer deliberately manipulated or created artificially, but stands as an intriguing slice of city life captured in camera-eye reportage.

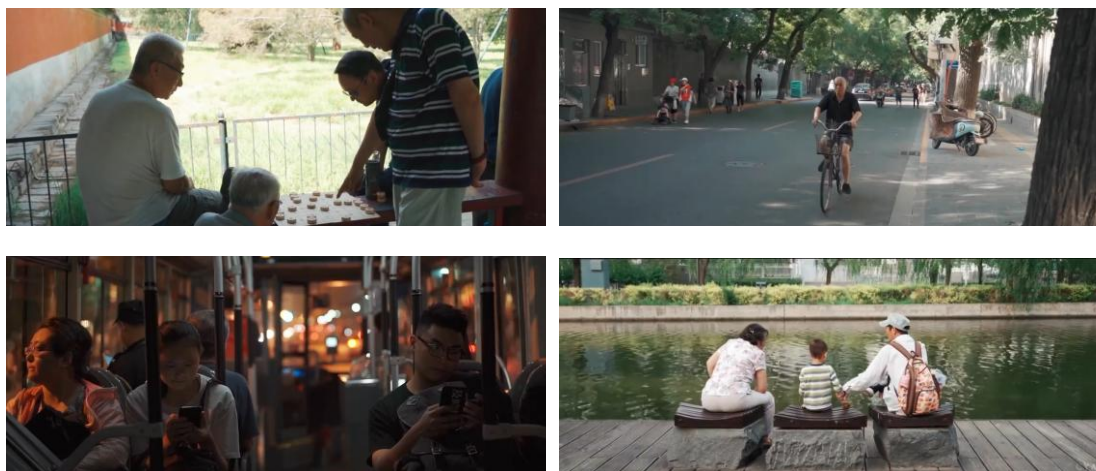


Fig. 7

Instead of stressing the depiction of specific characters, the creator intentionally applies a non-linear, non-narrative combination of group characters from all walks of

life to present the diversity and openness in urban life observations. Viewers can gain a holistic perception of everyday urban life through multiple character segments during viewing, breaking away from a single narrative thread seen in conventional videos. This collage-like presentation deepens the impression of Beijing's city imagery from multiple perspectives.

The visuals display unconventional urban portrayal since the creator chooses not to accentuate the city's modernity or traditional traces, or attach it with any stereotypical labels. Instead, characters' daily states have been filmed to articulate the warm-hearted life patterns in Beijing. These ordinary yet precious moments indeed express the city in a close-up and gentle manner. As suggested by the video title, "Listen to the city", the short video wholeheartedly invites the audience to listen to the city's daily pulses with attentiveness, framing the characters themselves as sources of these heart-stirring sounds.

The creator's character strategy prevents the city's image from being reduced to a stereotype or excessive packaging. Instead, it allows the audience to naturally grasp the lifestyle and atmosphere by diving into the everyday details of specific individuals during viewing. As a series of decentralized character shots unfold in the short video, the city image is endowed with relatable, down-to-earth qualities. Such character arrangement reflects the creator's deliberate rejection of top-down narrative manipulation. The absence of a central protagonist stimulates a multi-perspective viewing experience, suggesting that innumerable micro-experiences seeped in daily life jointly shape the ultimate city image. This approach reflects that the creators' city image construction is firmly grounded in real-life contexts.

In the short video "A French Chef Tours Xi'an", a French chef is defined as the central character, whose perspective as a newcomer to this city is employed to observe and experience Xi'an's cuisine culture. The chef character, as an active

participant in local culinary life, savors representative local dishes in a stepwise manner. The creator arranges the narrative around the chef's identity and behavioral path, offering a fresh viewing angle on everyday details often underestimated by locals (Li, 2025). The chef character, positioned within a cross-cultural context, proactively imparts layers of interpretive possibilities to the city's image, elevating the accessibility of Xi'an's daily life in the eyes of a broader audience.

By centering the narrative on this French chef, the creator deliberately draws comparisons between Biangbiang noodles (a Xi'an local cuisine) and Italian pasta, linking Xi'an's regional cuisine with globally recognizable food (see Fig. 8). The creator constructs a cross-cultural frame of reference, by giving full play to the character's cultural background and perceptual habits in the short video. This character strategy transforms local, ordinary elements into content to be understood and related to by audiences from diverse cultural contexts, lowering the threshold for engaging with local culture. The character thus helps connect locals and foreigners, successfully retaining the regional distinctiveness of Xi'an's daily life while boasting greater perceptibility and potential for wider dissemination.



Fig. 8

In this case, the creator visualizes the dining habits of Xi'an residents by designing the character's actions, such as displaying that Biangbiang noodles are served with side dishes and his dining process. Under this context, the creator transforms seemingly ordinary fragments of local culinary culture into accessible, engaging content with this character-mediated strategy. This also engages the audience to recognize Xi'an's culinary preferences in a relaxed and intuitive manner. As a result, city image construction goes beyond the mere accumulation of visual symbols and extends into the representation of its lifestyle and underlying cultural logic.

In the short video "Plane Trees, Western-style Houses, and Alleyways: A Glimpse into Old Times", the creator constructs an idealized version of Shanghai's daily life through purposeful character design. The character's narrative trajectory doesn't highlight the recreation of everyday realism, but links together symbolic locations such as historic villas, plane tree-lined streets, narrow alleyways, and banks along the Huangpu River.

This creates a tangible note of urban lifestyle, though the visuals weren't entirely shot from the routines of the ordinary. The video unfolds as a carefully designed scenario that reflects the rhythm and feel of life in Shanghai in sequential order. It starts from the character waking up in a historic villa, and then shifts to her strolls down the tree-shaded streets, and then to her leisurely coffee drinking across the river from Lujiazui while gazing at the skyline.

The character's walking route, choice of stops, and behavioral cues all revolve around the imagery building of a refined Shanghai lifestyle (see Fig. 9). The creator imbues everyday city life with layers of aesthetic and emotional meaning through carefully planned character movements and scenes. These visual settings taken from actual life are presented in a more aesthetic and stylized way. As guided by the character's actions and surroundings, viewers come to perceive the quality and pulse of urban

life, and attain clearer insights into cultural tones and meanings behind the city image.



Fig. 9

The creator constructs a semi-realistic version of daily city life by coordinating characters with spaces. The city image, shaped through this character strategy, doesn't feature the explicit depictions of daily patterns but creates a bridge between reality and imagination (Goldie, 2003). This approach characterizes Shanghai as a city imbued with poetry, subtlety, and a strong sense of humanistic warmth and cultural depth in the eyes of viewers.

Despite distinctive lifestyles and cultural characteristics in the three cities, creative approaches to representing daily city life through short video characters are similar in nature. In the cases of Beijing and Xi'an, creators use character-based strategies to portray aspects of life closely aligned with residents' lived experiences. In Shanghai's video case, stylized characters and meticulously orchestrated scenarios function to construct an idealized picture of daily life. They differ in the representational forms but share a common goal: stirring up viewers' intuitive understanding of local identity through characters' everyday actions, thereby cultivating more contextualized and human-centered city representations.

In terms of similarities, all three city cases follow a narrative that begins with characters' concrete actions and gradually transitions to a broader picture of urban patterns. Characters serve both as the portal of engagement and the key medium of delving into the urban life patterns, whether as the patchwork of everyday moments in

Beijing, the cross-cultural engagements in Xi'an, or the scenario-based citywide explorations in Shanghai. Such character-centered illustration of daily routines makes city images more emotionally intimate and supports audiences from diverse cultures with a valuable entry point into urban experiences.

Conveying the city's spirit

Representing the spirit of a city with character design is effective for visual storytelling. The identity, behavior, and narrative perspective of the character forge a visual pathway for expressing the city's essence. Compared to straightforward statements about urban characteristics, this character-enabled method is more engaging and easier to connect with emotionally, which enhances both the communicative power and emotional resonance in the content. As a result, it effectively strengthens the sense of affinity with the city's ethos. This section also conducts a comparative analysis of the three cities to explore contributions by this strategy to shaping images distinctive to each city.

In Beijing's short video "A Lucky Chinese New Year", Chinese shadow puppetry, a traditional intangible cultural heritage, threads the video as the central theme and a guide for the audience to perceive the city in depth through the portrayal of artisans. This video is far from simply showcasing the time-honored performance itself. It accentuates the labor and dedication of behind-the-scenes craftsmen whose ordinary actions of manipulating puppets in shadow play indeed embody a compelling sense of spiritual strength (see Fig. 10).



Fig. 10

The creator carefully presents traditional crafts such as puppet coloring and carving, using slowly progressing shots to highlight key, nuanced details of the artisan's work. These handcrafted techniques are rendered tangible and visible throughout the camera lens, in which artisan characters' dedication and patience stress their deep respect for the craft and commitment to the protection of intangible heritage. This measured and persistent approach goes beyond individuals' daily work but powerfully represents the values of cultural perseverance and inheritance that co-define the spirit of Beijing.

Behind-the-scenes operators as a whole are delineated as another kind of focal point in this short video. The creator carefully captures their precise control of lighting sources and puppet figures, presenting a scene of muted yet focused collaboration. Although these individuals do not function as the central characters who could secure the majority of attention, they are crucial to the smooth execution of shadow puppet performances. Their "invisible" presence mirrors the countless grassroots workers who contribute silently and selflessly to urban development. They have no desire to seek the spotlight, yet constitute the backbone of the city's cultural continuity and daily functioning. In short, the video conveys a quiet yet resilient city essence, allowing viewers to gradually understand the compelling strength embedded in Beijing's image through the plain yet genuine portrayal of these characters.

This character-centered approach offers an alternative to grand narratives or abstract symbols in articulating the city's spirit. It grounds the city's image in real-life actions

and the affective presence of individuals, thereby providing a more tangible and credible path to humanistic expression.

In Xi'an's short video "This is Chang'an", character design and visual composition are dually utilized to highlight the tranquil quality deep-seated in this historic city. The video unveils with modern urban scenes, showing dense traffic and vibrant nightlife to construct a diverse and dynamic image of the city. However, as the video progresses to its latter parts, it abruptly shifts to a serene scene where a female character drinks tea with contentment. Her calm and composed presence and slow movements stand in stark contrast to the earlier urban bustles. This rhythmic shift offers a moment of visual relief and an internal tension that helps voice the city's underlying ethos (see Fig. 11).



Fig. 11

The creator builds up the character around the symbolic act of tea-drinking that carries far-reaching cultural significance. In this shot sequence, the character doesn't navigate viewers visually through the city but embodies the city's spirit. She responds

to the complexity of urban life with her composed presence. By capturing the individual's subtle movements and crafting an immersive ambience, the creator lends a slow-paced, contemplative identity to Xi'an's core city image. Tea drinking functions as more than just a daily routine, but also the medium for individuals to embark on a deeper dialogue with the city. This directs the viewer away from the surface-level urban vibrancy towards a more profound comprehension of the city's spiritual essence.

The city ethos is conveyed naturally through detailed renderings of character behavior and everyday settings (Brennan & Pearlman, 2023). Characters serve as visual mediators and embody specific outlooks on life, guiding the audience toward perceiving the city's inner values in depth. This contextualized mode of presentation thus communicates Xi'an's urban character in a subtle yet persuasive way, adding emotional fabric to the city's image.

In the short video "Riding Sunset Rollercoaster on a Summer Evening, Watching Shanghai Dance a Tango", the character's actions, as a set of narrative threads, highlight Shanghai's inherent spirit of "living mindfully". The video characterizes a youngster's summer activities, following her journeys of leisurely listening to music while camping, appreciating the evening breeze on a sightseeing bus, and strolling through narrow alleyways in Shanghai. These seemingly trivial moments, when vibrantly presented through her presence and gestures, convey a sense of attentiveness and genuine engagement with daily patterns.

The key to this character strategy lies in the contextualized presentation of everyday behaviors. Viewers are guided into the character's way of life through a relaxed narrative rhythm and depiction of her natural, unforced moments alone. The character loses herself in music with eyes shut amid the tender breeze, or wanders through alleyways bathed in dappled light (see Fig. 12). These simple gestures reflect the

character's pursuit of a higher quality of living while embodying Shanghai's urban refinement.



Fig. 12

Additionally, the character's engrossed presence in various everyday scenes expresses the concept of "living mindfully". The creator turns Shanghai's urban ethos from an abstract concept into a perceivable lifestyle by tightly fusing the character's state in life with the city space. This reflects the city's modern charms and subtlety deep-seated in its cultural depth.

As enabled by the zeal for life conveyed by the character, Shanghai's constructed city image in the short video transcends functional or stereotyped expressions. Instead, it presents a distinct city spirit shaped through an interplay of the character's actions and emotions, enabling viewers to connect with and grasp Shanghai's spiritual core in greater depth.

The short video cases for the above three cities are distinguishable in the portrayal of urban spirit. These distinctions stem primarily from the creators' differentiated emphases in character strategies, ranging from occupational roles to cultural backgrounds or behavioral states. These different approaches lead to divergent paths in city image construction. Moreover, these character strategies are not purely driven by personal creative choices. Rather, it vividly showcases the structural differences shaped by each city's social fabric, historical context, and cultural temperament.

The subtle tendencies shown in character design and behavior arrangement often suggest the continuation of a city's inherent cultural values. As a result, depicting the city ethos through character portrayals doesn't just reflect individual traits, but also visually represents the city's broader identity and spiritual values. This study thus identifies that such a process underlines short videos' capacity in perceptively capturing urban identity, rearticulating local spirit and distinctly manifesting each city's spiritual qualities, thus contributing to the construction of recognizable, differentiated city images.

Although their spiritual connotations differ, these city-based short videos reveal a shared construction mechanism in the representational approach. In detail, the characters' speech, behaviors, and interactions with daily life are infused with symbolic meaning, transforming intangible spiritual values into visible, lived experiences. This character-centered approach allows city image construction to stretch into values and cultural identity, thereby strengthening the emotional connection between the city and its audience. This construction rationale solidifies the cultural foundation of city images and provides visual support for the distinct recognition of each city's ethos.

Evoking reflections on cities

When constructing city images, characters not just hold extrinsic representations but also facilitate audiences' reflection on the city. Characters, when placed in specific urban contexts, normally lead viewers to reinterpret the symbolism of the city or reassess the relationship between individuals and their urban surroundings. Such purposes are fulfilled as their actions, language, or personal backgrounds are unfolded in the short videos. In the short video format, beyond merely showcasing visual aesthetics or daily patterns, characters evoke emotional resonance and shared values,

leading audiences into a more reflective viewing mode (Cohen, 2013).

The circumstances, emotions, and behaviors embodied by characters often point to the in-depth social structures and emotional dynamics of the city. These tensions within character portrayals enable creators to uncover once-overlooked aspects of urban life and transcend from superficial presentation of characters toward insights into their deeper thinking on each city. The struggles, alienation, or dilemmas faced by characters become annotations of urban reality and offer both emotional and narrative anchors for reflective expressions of the city (Meng, 2024).

In conclusion, this section delves into the theme of character-initiated reflections of the city, exploring how characters are depicted to steer the audience toward a profound re-interpretation and re-evaluation of the city itself and individual experiences within it.

Reflections on the city

Amid rapid urbanization, characters have grown as mediators for short video creators to engage with urban issues and express value judgments. This elevates city image construction from surface-level representation to more contemplative and in-depth means of expression. This section focuses on how characters assume the role of illustrating reflections on cities, thereby examining how character design conveys creators' observations and ruminations on urban transformation.

In the short video "Beijing Citywalk: Life is Not Elsewhere", the creator ponders over the question, "What kind of city can truly be a national capital?". Instead of adopting abstract commentary or direct evaluations, the video unfurls a rich display of characters to navigate the audience into reflecting on this question. The final scene is characterized by the Chinese phrase "包容" (which means inclusiveness) as the

thematic conclusion. However, this term isn't reducible to a simple label, but emerges naturally from the preceding character compositions, whose diversity and coexistence embody the very idea of inclusiveness.

The creator chooses to incorporate a diverse range of cultural figures, including Peking opera performers, crosstalk comedians, and pop singers (see Fig. 13) into the short video. These individual characters differ in their performance styles and represent art genres of traditional opera, folk art, and contemporary popular culture, respectively. Their alternating appearances in the video demonstrate Beijing's cultural inclusiveness and coexistence. Each performer's distinctive cultural background and expressive state collectively construct a vision for the capital city: embracing an extensive spectrum of traditions and cultural influences.



Fig. 13

In addition to these art performers, the video also features an elderly man strolling through a hutong, enthusiastic young concertgoers, and young people enjoying the city's nightlife (see Fig. 14). These characters lead varied lifestyles and daily rhythms due to their varied social backgrounds. Their appearances highlight the diverse social structure and ways of living accommodated by the city. The creator blurs the individual differences but leverages the presence of characters to underscore the philosophy where a city's ability to embrace such everyday variety is essentially a meaningful characteristic.



Fig. 14

In brief, Beijing’s city image is turned from a mere depiction of characters into a wider reflective thinking on public life and space. This reflection-grounded approach to image construction enables the “capital city” to transform from simply functioning as an abstract political concept to taking on concrete, tangible social connotations.

In the short video “My View of Xi’an”, a returnee back in Xi’an is presented as the main character whose perspective leads viewers to explore the city’s transformation in depth. Such character design shapes a unique standpoint for observing urban development and historical evolution. This character stops and stays in front of the ancient city walls, touches its weathered bricks, and expresses his heartfelt reverence at the profound weight of millennia-old history. He then enters the city’s central business district (CBD), where the towering buildings and hustling traffic flow arouse in him a keen awareness of modern development (see Fig. 15).

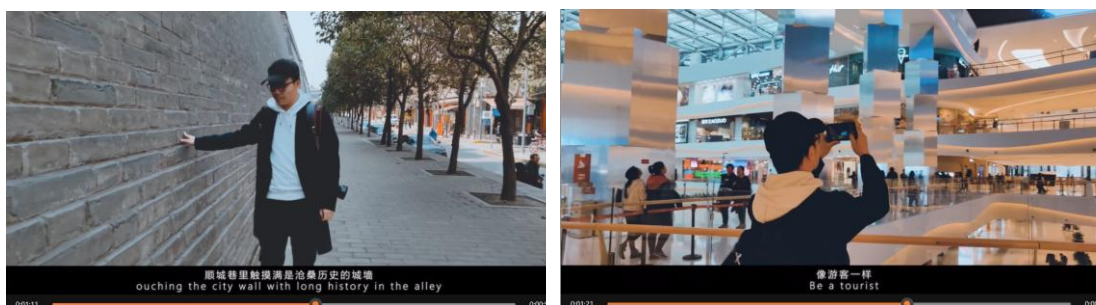


Fig.15

This refreshing narrative structure helps the city image to shift from a wide-ranging visual presentation to a more emotionally charged personal experience. As viewers

follow the character's transitioning emotions and facial expressions, they are explicitly presented with the city's evolving nature over the course of time. His deep reflections on the contrast between the once-mysterious Chang'an and today's increasingly commercialized internet-famous Xi'an both highlight the changing facets of urban communication and the redefinition and reinterpretation of local culture in answer to media transformation requirements. This shift represents a physical spatial change in the city and signals a reconfiguration of cultural understanding and identity for inhabitants within it.

This short video renders the urban transformation as a process to be tangibly sensed by ordinary people through the character's plain yet sincere language and his sensory experiences. This reflection emerges as an emotional insight derived from everyday life. As the character reacts to the consequences of urban development with a mix of awe and nostalgia, he assumes a dual role of observer and narrator, thereby upgrading the city image construction into an in-depth dialogue with a sense of belonging, memory, and change.

This short video, framing the narrative around the returnee, imparts Xi'an's urban development with a personalized representation. As the character travels across this strange yet familiar city, his observations and emotions lead to his renewed understanding of the so-called "hometown", offering a more humanized perspective to dive into the city's ongoing transformation.

In Shanghai's short video "Falling in Love with the Unique Blend of Prosperity and Civility in the Magic City", the creator applies a character-driven approach to re-examine and reconstruct the city's temperament. Conventional media portrayals tend to evoke a fixed set of associations and stereotypes, such as modernity, fashion, and a perception of speediness, by referring to the term "Magic City". However, the creator avoids adopting these label-like definitions, but draws on the character's individual

experiences and everyday perceptions to answer the questions proposed in the video and redefine the genuine temperament of Shanghai.

The video begins by introducing a character positioned within a typical high-pressure metropolis setting. The character, surrounded by heavy traffic and wearing an exhausted expression, reflects her stressful life caught in the relentless pulses of the city (see Fig. 16). Rather than romanticizing or artistically embellishing this living state, the creator voices repulsion for such a highly routinized urban lifestyle through the lens of the character’s internal monologue and emotional shift. This act of rejection, by its nature, becomes a reflection on Shanghai’s widely recognized image of “Magic City”.

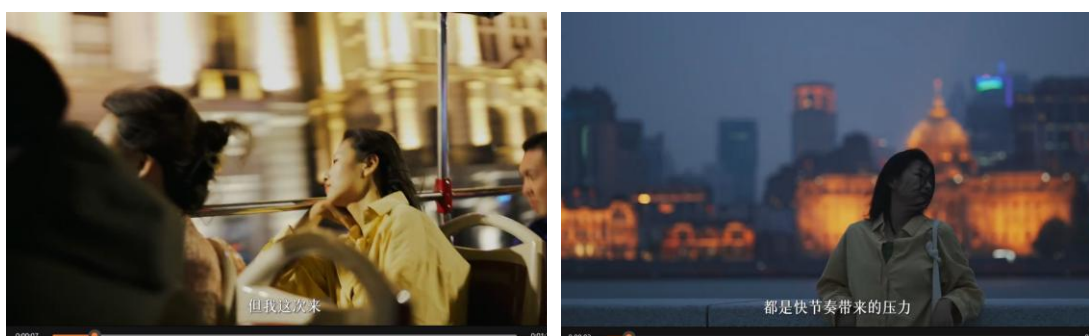


Fig.16 (a)

As scenes shift, the character gradually changes her behavior by initiating the reconstruction of “Magic City temperament”. Adopting the dual-structure approach to two contrasting city image expressions reveals what the creator deems as Shanghai’s genuine temperament. In one aspect, the character is filmed standing amid the city’s nighttime skylines and neon lights, highlighting the material prosperity offered by the city. In another aspect, the character wearing vintage attire savors tea and tries painting, unveiling Shanghai’s more refined and cultured dimension.

These slow-paced yet focused moments display an appreciation for cultural refinement and a rich intrinsic world. The character’s rhythm, posture, and

surroundings collectively shape a daily urban life that is aesthetically diverse and emotionally engrossed, thereby challenging the stereotype of a singular imagery of Shanghai.

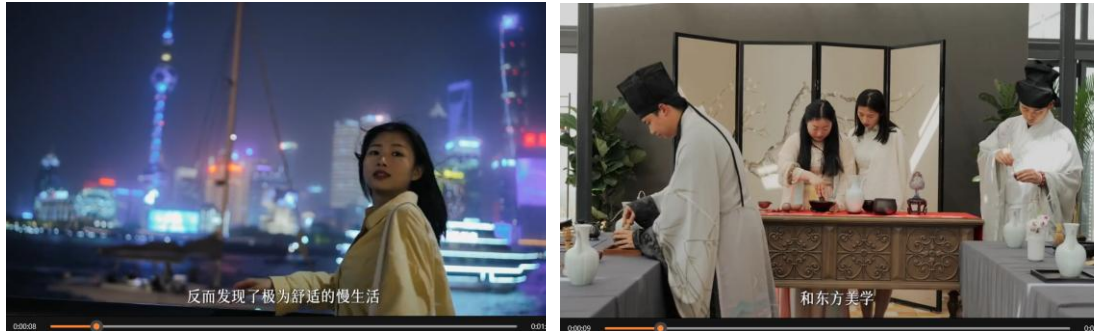


Fig. 16 (b)

This effective character-driven strategy redefines Shanghai into a confluence of prosperity and elegance. The character no longer keeps up with urban rhythms blindly and passively, but instead consciously reflects on and opts for her lifestyle as an individual. Consequently, the creator offers a reflective response to the question of Shanghai’s urban temperament through the character’s exclusive perspective. Therefore, the city image experiences substantial reconstruction by breaking free from the visual intensity of skyscrapers and neon lights that signal hectic vibes, and cultivating a more humanistic and culturally nuanced representation.

Points of focus and modes of expression differ significantly despite the shared character-enabled city reflections in the three city cases. Beijing’s video case uses a mosaic of individuals from various cultural and social backgrounds as characters, intertwining them within the same urban narrative to identify the meaning of public space underlying Beijing’s identity as a “national capital”.

Xi’an’s video case combines the city’s development with personal emotional experience through the returnee character, turning urban transformation into a perceptible memory reconstruction (Geng et al., 2020).

Shanghai's video case redefines the so-called temperament for a "Magic City" by following in the footsteps of an individual character who keeps challenging stereotypical attitudes and broadening the perception of urban identity beyond speed and efficiency towards refined aesthetics and cultural taste. These differences reflect the creators' keen observations of each city's exclusive qualities and creative efforts to express them with differentiated strategies. As a result, the short video creation suggests that city-specific reflection pathways are closely tied to the narrative standpoint introduced by each character.

As for similarities in underlying mechanisms, each video case avoids applying direct commentaries on the city and instead relies on characters' concrete actions, situations in life, and emotional states to direct the audience toward engaging with the urban context in further depth. Characters, beyond their information communication role, are tasked with initiating reflections and expressing values, whose perception, judgment, and emotional responses to urban development work together to visually construct the city as a space of thoughts and resonance. This character-driven reflective pathway expands the expressive boundaries of city representation and enables more textured emotional relatability.

Reflections on individuals

Following the continuous evolution in cities and their spatial orders, characters have emerged as the key entries for creators to explore the relationship between individuals and the city. Although these characters don't explicitly depict the city itself, their presence often reflects the living conditions of certain groups of people within the urban environment. This section takes a deep dive into the individual-level reflections proposed by short video characters, analyzing their layered significance in city image construction.

Beijing's short video "You Have to Climb the Great Wall at Least Once", while documenting the creator's ascending experience of the Great Wall, captures an elderly man peddling goods in its scenic area through camera lens shifts. The man, as a highly symbolic figure, hangs up his merchandise on the tree branch (see Fig. 17). Despite the apparent hardships and his crude working environment, he wears a consistent smile, greets and bids farewell to tourists with wholehearted warmth. His gestures and demeanor convey calmness and contentment, presenting an individual attitude in stark contrast with the fast-paced rhythm of urban development.

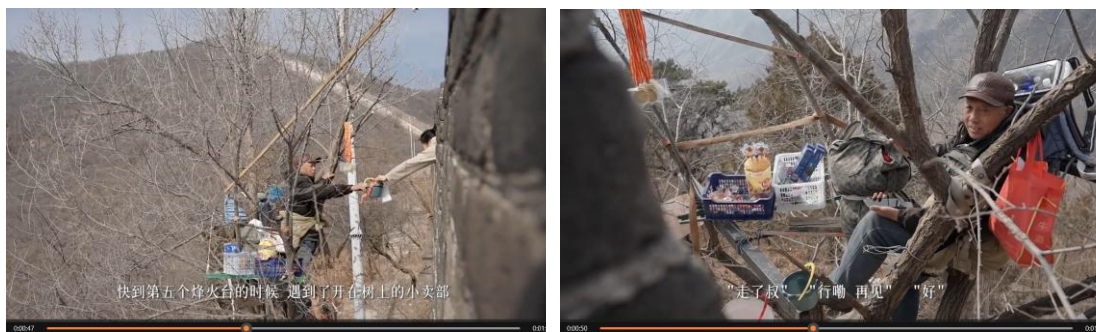


Fig. 17

This enlivened character portrayal drives deeper reflection on individual circumstances within urban evolutions. The elderly male character, as a grassroots laborer in his senior age, reveals through his presence that those marginalized by dominant narratives obscurely contribute to the city to support its smooth daily functions in their own way. His modest means of livelihood underscore the "invisible" status of marginalized individuals from unseen urban corners. Meanwhile, his enduring optimism endows this unobtrusive occupation with a tranquil resilience and an unassuming sense of value (Plantinga, 2008).

It is to be noted that the creator includes no direct judgment on the character's living condition or fate, but employs visual contrasts and controlled pacing to guide the video viewers to ponder on the neglected fragments of life against the rapid

urbanization backdrop. This reflection not only highlights the unequal relationship between individuals and the city, but also facilitates viewers to reconsider the definitions of “success” and “value”. Such character portrayal, as a key reflection mechanism initiated by short video characters, constructs a gentle yet compelling critical perspective in the short video, leading the viewers to focus on and pay respect to tangible lived experiences beyond the overarching urban narrative. It also manifests the growing trend of contemporary city image construction toward diversity and human-centered values.

In Xi’an’s short video “The Only City Worthy of the Glory of the Tang Dynasty is Chang’an”, historical imagery of “Chang’an” threads the overall narrative, building up a character-driven storyline abundant with cultural depth. The creator intentionally selects two highly symbolic local figures, Xuanzang and Xu Wei, to spark reflections on individuals’ living and spiritual states (see Fig. 18).



Fig. 18

The first half of the video portrays the historical Buddhist monk Xuanzang, whose identity is used to unfold a poetic reflection on spiritual belief and the state of life. When the character says, “All beings are bound by joy and sorrow, yet I wish to see all the flowers of Chang’an with you in a single day”, he transcends himself from a mere religious symbol to a spiritual emblem that goes beyond history and articulates inner struggles. His imagery against the bustling city context propels further pursuits of inner freedom, meanings in life, and individual circumstances.

The creator then introduces the contemporary rock musician Xu Wei as a modern embodiment of Xi'an's ethos. In contrast to Xuanzang's introspection and compassion, Xu Wei's presence brings a more modern and emotionally charged reflection. His well-known statement, "Don't grow numb with age—burn with passion", directly addresses the spiritual fatigue and confusion experienced by today's countless individuals. The representation of his desire to break free from everyday rigidity calls for emotional authenticity and vitality beneath the city surface.

Juxtaposing symbolic figures from different eras facilitates the creators' construction of individuals' reflection pathways that blend tradition with modernity. The city is sublimed from a microcosm of visual glory back in the Tang dynasty into a space that continuously prompts self-inquiry and reflection on personal values. In this video, the characters help with emotional expression and intellectual transformation, upgrading the city's image into a far-reaching contemplation of attitudes to life and existential meaning (Kaul, 2014).

In the short video "Where Were You This Winter? How Was It?", the first-person narrative enables reflective thinking on personal development and change. The character is simply illustrated as an ordinary youth assigned with no specific professional or social label. The video captures her embodiment of temporal flows and intrinsic transformation. The depiction of personal development intertwines change and continuity. Specifically, the character's presence in Shanghai remains unchanged, while she sees changes in herself as she has her hair cut short, experiences stronger mental fatigue, and takes steps closer to realizing her dream. This reflection both recounts the individual's life path in Shanghai and represents her state of coexistence with the city.

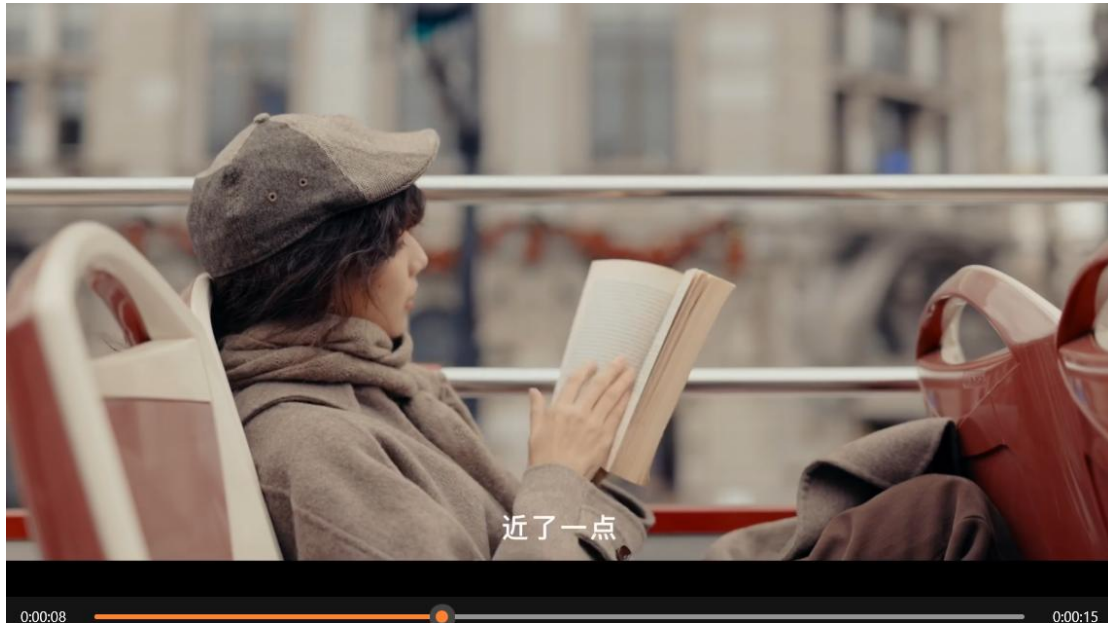


Fig. 19

Nuanced changes found in the character herself are depicted in a calm and steady narrative rhythm to invite the audience to probe into the challenges and emotional anchors of individuals residing in metropolises like Shanghai. This reflection, grounded in personal experience, positions “personal development” as the thread linking the individual with the city. The character’s emotions, derived from the city context, are amplified within the temporal setting of a winter in Shanghai. Winter, as the most tranquil time of the year, naturally arouses emotional retrospection and carries the weight of time and a respite from the bustling life rhythms. Thus, the creator augments the character’s emotions and infuses the personal development theme with a tender warmth amid the chilliness and stillness in winter.

This character strategy transforms the character’s changes into more than just a record of physical or everyday experiences by endowing a sense of temporal and spatial resonance with broader meanings. Winter turns the city from a mere spatial container into a contextual structure that bears and absorbs personal experiences due to its ability to emotionally connect the city and individuals. The character’s mental fatigue, persistence, and hope within the chilly season of winter mirror Shanghai’s ongoing

evolutions, fostering a deeper comprehension of individual circumstances within a macro urban environment.

The short video cases from these three cities share the expressive rationale of how characters elicit reflective narratives. The characters are central to conveying further textured meanings, as seen from cases of the elderly man selling souvenir goods in Beijing's tourist attraction, the juxtaposition of historical and contemporary figures in Xi'an, or the self-narrated development journey of a young woman in Shanghai. These portrayals lead viewers to engage with the individual experiences within urban life. City image undertakes a more humanized and tangible (Felleman, 2014) facet through the character-enabled representation of personal living states. Characters thus serve as important gateways for reflecting on urban systems and individual values, adding intellectual depth to the city representation in short video cases.

At the same time, each city's short video case differs in its reflection themes and modes of representation. Beijing's video highlights the overlooked status and value of individuals peripheralized in urbanization. Xi'an's video explores the continuity and transformation of individuals' spiritual core within cultural transmission. Shanghai's video case centers on the personal development of modern youth. These differences, despite originating from the creators' character strategies, also reflect fundamental distinctions in cultural context, historical heritage, and social structure among Beijing, Shanghai, and Xi'an. This character-centered reflective approach indeed sublimes city image construction from an illustration of visual landscapes to in-depth dives into the lived experiences and inner worlds of urban residents. To summarize, each city is portrayed in a more authentic and nuanced manner thanks to the character-empowered short video representations.

Conclusion

This chapter thoroughly analyzes the distinctive functions of characters in city image construction and identifies the similarities and differences in the creators' character strategies applied to these three cities.

All characters perform three major functions in short videos. First, they operate as narrative threads that connect urban spaces and enhance the structural and rhythmic coherence in visuals. Second, they, as facilitators of everyday life, present urban routines and help viewers perceive the city's atmosphere and lifestyle. Third, they work as emotional triggers to navigate the audience towards reflecting on the city's cultural temperament or individuals' living circumstances. This character-driven approach transforms the city from a symbolic backdrop into a tangible and emotionally relatable cultural space. Characters thus become key instruments for creators to reinforce narrative structure, guide viewer engagement, and convey the city's ethos in short videos.

However, notable distinctions in character strategies across the three cities exist behind these structural similarities. In Beijing's video cases, character design tends to fuse cultural depth with real-life concerns, often portraying group characters, cultural figures, or overlooked grassroots individuals to construct inclusiveness and warmth. In Xi'an's video cases, the collocation of contemporary and historical characters emphasizes the city's cultural diversity built upon the harmonious coexistence of traditions and modern allure. Such characters often serve as temporal guides that reveal the historical context and spiritual legacy embodied by the city. In Shanghai's video cases, character strategies stress more on individual expression and aestheticized versions of everyday patterns, outlining personal experiences and emotional states to represent the city's modernity, openness, and cultural sophistication.

These distinctions dually reflect the fundamental variations in the three cities' cultural

contexts and social structures, and the highly adaptive and place-specific nature of character strategies. Characters not only enrich the expressive approaches of short videos but also facilitate the underlying rationale behind city image construction to shift from undertaking spatial representations to embodying subjective experience, and from seeking extrinsic observations to adopting intrinsic reflections. Finally, this signals a broader shift in contemporary city image construction, which increasingly values individual perception and cultural interpretation.

Chapter 7 The Role of Post-production in Constructing City Images

Introduction

Following the rising popularity of short video platforms and leapfrog advancements in digital technology, post-production has come into common use as a creative tool for content creators. The masterful use of post-production techniques bestows new dynamism to city image construction, enabling a greater expressiveness and appeal in image representation within a limited timeframe of short videos. Additionally, post-production is integral to organizing and enhancing visual symbols, and strengthens the expressive power of city imagery through creators' efficient control of editing rhythm, meticulous arrangement of shot combination, and appropriate addition of visual special effects.

This chapter is divided into three aspects, namely editing rhythm, shot combination, and visual special effects, aiming to unveil their respective applications in city image construction in the three city cases. It includes a city-specific comparative analysis that examines how creators in each city emphasize local features and construct the ultimate city images with post-production techniques. This chapter also aims to uncover the functional mechanisms of post-production in shaping city imagery and provide further insights into the employment of state-of-the-art technologies in image construction.

Editing rhythm in city image construction

Editing rhythm, as a key element to post-production, is defined as the speed and variation of scene transitions within visual content. It influences the viewing experience and determines the narrative flow and emotional tone in storytelling. According to Bordwell, Thompson, and Smith (2004), editing rhythm in feature-

length films helps shape the temporal flow and spatial continuity that connect, transition, and hold camera shots. For example, a fast-paced editing rhythm enhances tension and dynamic energy in action films, whereas a slower rhythm is often found in lyrical or documentary films to prolong the audience's perception of time in crucial moments. In the latter case, people can take the time to feel changes in scenes and emotions carried by the characters (Pearlman, 2012).

Short videos, especially those shared on social media platforms, often feature a brief duration and densely packed information. Specifically, the proper use of editing rhythm in the short video context can efficiently convey a city's spatial structure, cultural ambience, and emotional tone (Pearlman, 2012) for a better construction of city images. Editing rhythm can also represent the diverse and complex sensory experience of city places through contrasts between varied paces.

Editing rhythm applied to short videos fulfills three functions. First, it shapes the dynamic presentation of the city image by accentuating energy and modernity through quicker paces, and conveying historical and cultural depth with slower paces. Second, it breaks through spatial barriers with rapid scene changes. As noted by Cutting, Brunick, DeLong, Iricinschi, and Candan (2011), swift shifts, when applied to short videos, enhance spatial integration in film productions and enable viewers' continuous and layered perception of the city. Third, it compresses or extends the sense of time, unfolding distinctive city narratives within a short viewing period and instantly engaging viewers.

Editing rhythm and sensory experience of city spaces

Editing rhythm helps present the sensory experience of city places. Such experience includes the flow and transformation of physical space, as well as the cultural atmosphere, social activities, and emotional representations within the city. Pacing

variations and smooth connection of shots lead viewers to sense the city's temporal rhythms, spatial characteristics, and overall temperament (Dancyger, 2018). Based on this concept, this study has identified that fusing editing rhythms with the representational means of places helps construct an image distinctive to each city.

The following sections of this chapter analyze short video cases in Beijing, Shanghai, and Xi'an, exploring the varying representation of city-based sensory experience and contributions to image construction by editing rhythms.

In short videos about Beijing, creators apply various editing rhythms to diverse scenes to sensorily represent the city's multiple facets. The editing rhythms of scenes with time-honored landmarks stand in marked contrast with those depicting the city's modern metropolitan imagery, reflecting the tension between historical continuity and flourishing contemporary development. Fast-paced editing creates an efficient, compact visual rhythm and thus constructs the city as a lively international metropolis, while slower-paced editing guides viewers into the city's historical texture and cultural warmth, deepening their perception of Beijing's cultural profundity.

For example, the short video "Will Your First Stop in 2024 Be the Capital Beijing?" represents the vigor of Beijing's modern commercial districts (see Fig. 1) via fast-paced editing. The video stitches together a series of short shots filming the hustles and bustles of shopping streets, financial districts lined with high-rises, and crowds, thus building up a tight and dynamic visual flow. Beijing's impressive identity as an economic hub and a global city is underscored by the high-density visual information and a continuous sense of movement underlined by the creator. This editing strategy enhances the viewers' immersion in a short timeframe, since rapid shot transitions constantly stimulate visual perception to produce a rhythmic viewing experience, and help viewers forge direct, affective bonds with the city image.

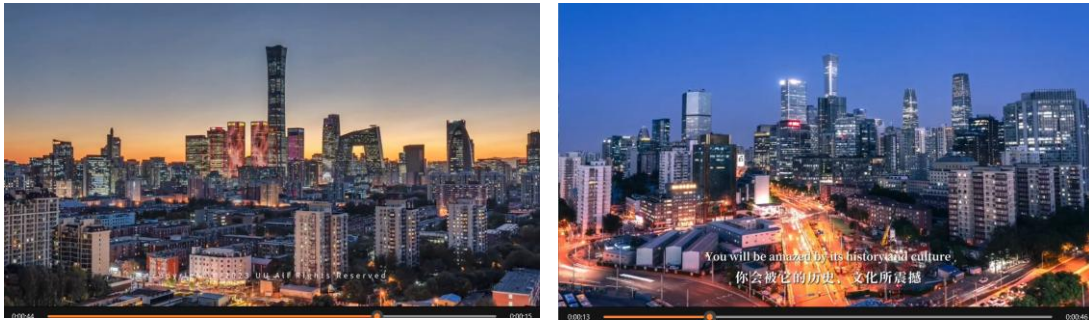


Fig. 1

In sharp contrast, the video “The Most Characteristic Beijing” features a much slower editing rhythm to emphasize the everyday texture of hutong culture. The creator employs long takes and low-frequency cuts to allow for more pauses and transitions within each shot. In this way, viewers can examine details of traditional courtyard residences, small-sized street shops, and locals’ daily patterns (see Fig. 2). The slower pacing enhances the immersive quality in the spatial experience and subtly extends the perception of historical time flows and cultural memory. As a result, Beijing’s historical atmosphere is conveyed gently and thoughtfully.



Fig. 2

The contrast between editing rhythm paces reflects differences in types of location, creating a shifting pathway for viewing perceptions (Smith & Henderson, 2023). The variety of editing rhythms in short video production contributes to multi-layered image construction. Slow rhythms augment the historical depth and cultural warmth. Fast rhythms communicate efficiency, prosperity, and the tempos of a modern metropolis. By alternating between and juxtaposing rhythm paces, different temporal

layers and spatial textures are interwoven. This strategy smoothly shapes Beijing as a city both firmly based in history and embedded with a forward-looking vision.

In short videos about Xi'an, slow-paced editing highlights the calm and depth of this well-recognized historic city, reinforcing its sensory impression as a timeless tapestry of history and culture. In contrast, fast-paced editing brings in a vibe of motion and momentum, displaying the city's daily pulses and diverse urban character. The shift between and a blend of these editing rhythms both allow viewers to effortlessly switch between the past and the present, portraying a multi-layered and more enlivened perception of Xi'an.

For example, the short video "Cinematic Travel Short of Xi'an: I Want to Accompany You from the Morning Bell to the Evening Drum" is structured in progressively ascending editing rhythms (see Fig. 3). The video begins with slow-paced cuts that unfold historical imagery such as ancient sculptures, city walls, bell towers, and gates. On-screen time is extended in these scenes, drawing viewers into Xi'an's solemn and tranquil ambience and cultivating in them a rich sensory recognition of its cultural depth. As the video progresses, editing rhythms speed up, cutting to vibrant markets, street activity, and city nightscapes. This transition creates a stark contrast in rhythm and emphasizes the visual shift from stillness to movement, showcasing multi-faceted ways of presenting city images across different times of day and narrative contexts.

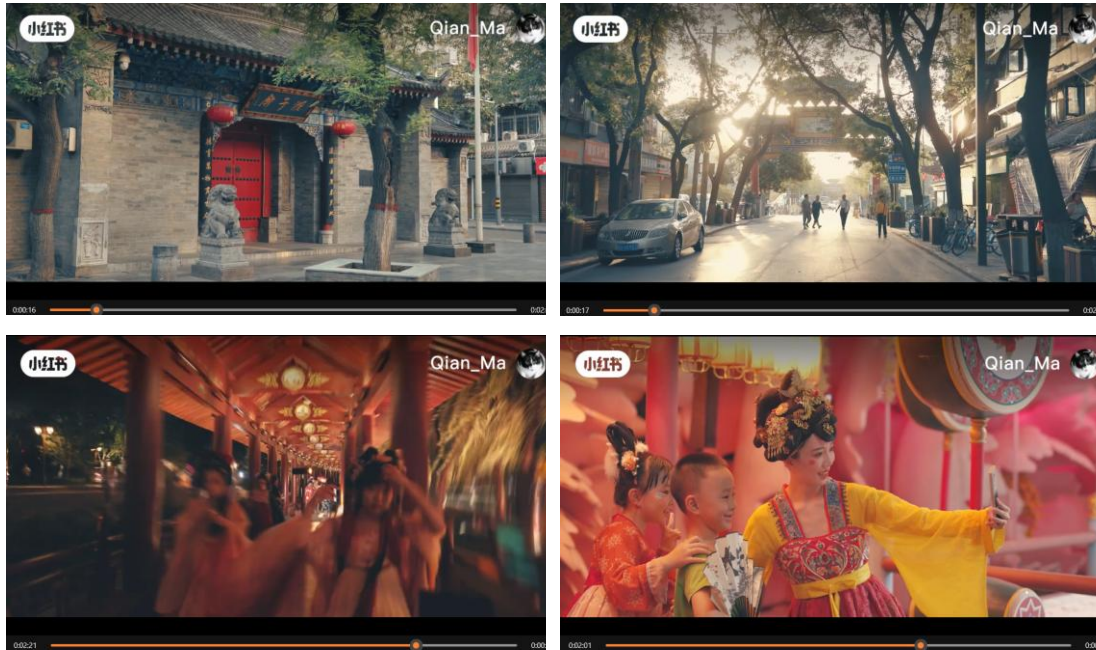


Fig. 3

Great Tang All Day Mall is portrayed in a particularly notable video segment. The creator grasps the flow of crowds, lighting, and interactive scenes at the night market by adopting rapid shot transitions and compactly arranged visuals. This shift in editing rhythms transforms the once-distant sense of history into a living part of the present, which coexists and interplays with modern life, visually imbuing a festive and energetic feel to Xi'an's nightlife. The ascending editing rhythms lead to a shift from serene reflections to lively celebration.

The rhythm control in this Xi'an-based short video serves as a narrative strategy for creators. It transforms the city image from a static historical scene to a dynamic vessel of everyday life, revealing the regenerative power of thriving culture in a modern urban context. The interplay of fast and slow rhythms adds emotional tension to Xi'an's visual representation, which leads viewers to engross themselves in both historical profundity and the vitality of daily city life. This rhythm-driven viewing experience ultimately supports multi-layered city image construction on both sensory and emotional levels.

In short videos about Shanghai, creators often combine fast-paced editing with aerial shots to present the city's nightscape. These dynamic sequences, when paired with upbeat background music, reinforce the city's international metropolis image. Creators, through rhythmic arrangements of city space transitions, guide viewers to sensorily experience the city's prosperity and sense of order, concretizing its urban rhythm.

To illustrate, the short video “Shanghai Impressions (Aerial) — The Glamorous Night View of the Bund” presents the Bund's core areas in quick succession within just a few seconds via fast-paced editing. Scenes covered in the video include the financial district with a forest of skyscrapers, the bustling arterial roads, and the well-illuminated riverside buildings at night (see Fig. 4). The aerial shots move rapidly across the sky, shifting perspectives through overhead shots and sweeping transitions. This nurtures a strong visual rhythm by continuously breaking limitations of a single viewpoint. The editing strategy emphasizes the city's perception of rapidity and intensity, through which the audience can grasp the high-density spatial structure and dynamic pulses of Shanghai nightscapes in a short duration.

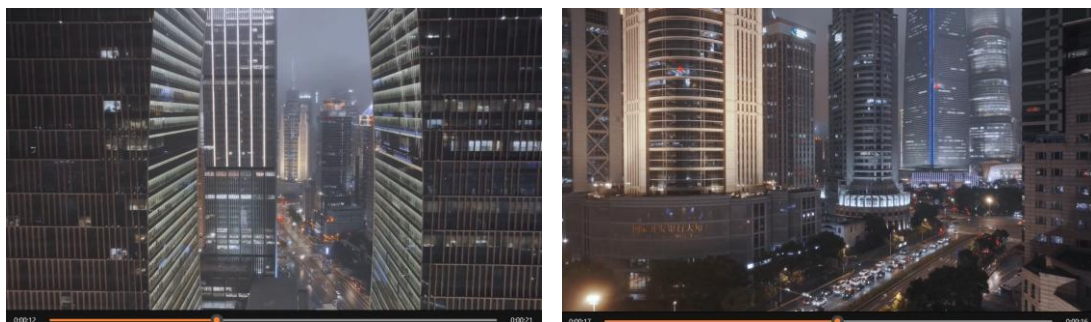


Fig. 4

The creator intentionally inserts several slower-paced shots in the middle of the video, such as water rippling across the river surface and reflections of city lights (see Fig. 5). These shots feature noticeably longer intervals between each cut and a more expansive visual feel, offering viewers a brief emotional pause upon experiencing the

intense visual impact. This variation in rhythm not only makes overall pacing more layered but also diversifies the sensory experience of the city space, presenting viewers with moments of stillness within the motion.

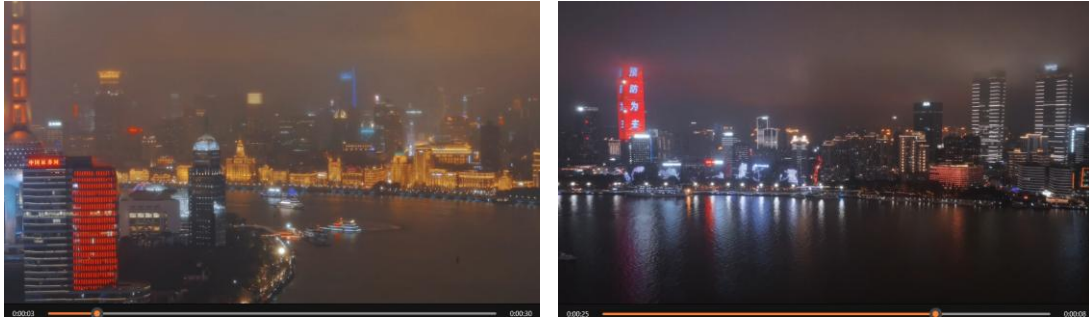


Fig. 5

Beyond the simple representation of city space, editing rhythms in this short video function as strategic tools that guide viewers' perception. A faster editing rhythm highlights the city's flourishing and efficient qualities, while the slower pace allows viewers to capture the finer details. These rhythm alternations construct a spatial experience blending movement with stillness and balancing intensity with tranquility in Shanghai.

Overall, editing rhythms applied by creators to short video cases of Beijing, Xi'an, and Shanghai share a certain pattern: alternating between fast and slow pacing in different scenes across city spaces to enhance the content's visual flow and expressiveness. Slower pacing finds its wider usage in representing historical culture or emotional ambience, while faster pacing lays greater stress on the energy and efficiency of modern city routines. This editing strategy strengthens the narrative tension of the city imagery, thereby deepening the viewer's immersion. Thus, editing rhythm, as a regulation mechanism for sensory perception and emotions, is key to shaping the sensorial sensations of city images.

Nonetheless, each city presents a distinct perceptual approach to its editing rhythm. In

Beijing's short video case, the juxtaposition of fast and slow pacing outlines spatial contrasts, enabling the quick visual switch between towering skyscrapers and deep hutong alleys and bringing forth the city's multilayered structure and spatial tension.

The creator of Xi'an's short video employs the gradual ascent of rhythm from stillness to movement, navigating viewers from appreciating engrossing historical scenes to present-day urban everyday details, and thus imbuing the city's spatial experience with a significant temporal depth.

In Shanghai's video case, a consistently speedy editing rhythm uncovers the city's fast-paced lifestyle and urban density. Meanwhile, the insertion of a few slower-paced shots slightly disrupts the rhythmic uniformity but nurtures peacefulness within the vigorous motion. This produces a visually balanced sensory rhythm that merges intensive energy with moments of pause.

This video case proves that the skillful control of editing rhythms can facilitate creators to elicit sensory resonance in viewers within a limited timeframe, reinforcing both the visual impression and emotional vibes across spaces in a city.

Editing rhythm and transformation of city scenes

Besides the above roles, editing rhythm is also key to orchestrating city scenes and driving spatial storytelling. Variations in pacing, shifts between video segments, and shot connections are used by creators to arrange shooting content across places, periods, and atmospheres, imparting a fluid, orderly visual structure to the city image. Editing rhythm technically supports smooth shot connection during scene transitions, as identified by Magliano and Zacks (2011), and serves as a creative strategy that navigates the viewing journey and shapes the expressive structure of the city imagery. Rhythmic design empowers the camera lens to switch among and connect different

city scenes in a short time, highlighting the intrinsic diversity and dynamism of city spaces. As a result, the city image gradually unfolds through an uninterrupted experience of time and place.

The following section examines the roles of editing rhythm in scene transitions, focusing on specific techniques and creative strategies in short videos centered on Beijing, Xi'an, and Shanghai.

In the short video “Winter at Shichahai: Don’t Miss the Unique New Year Atmosphere in Beijing”, editing rhythm is vital to coordinating scene shifts, bridging the construction of festal atmosphere and representation of spatial diversity. The video was filmed in a well-lit and lively pace, linking representative celebratory venues during Chinese New Year, such as Nanluoguxiang (South Luogu Lane), Yandai Xiejie Street, and Shichahai (known as a historic area in Beijing), with high-frequency shot changes (see Fig. 6). This creates a compact and visual rhythm that feels both flowing and compact. The continuity of rhythm provides a clear logic for the transitions between city spaces.

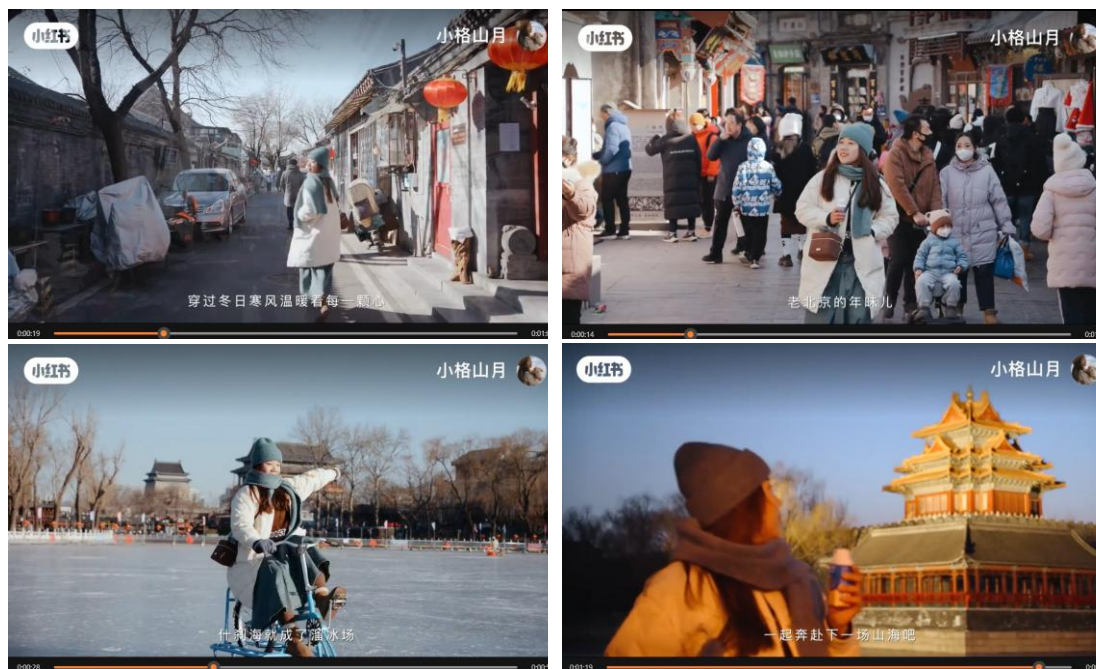


Fig. 6

The creator intentionally combines tracking shots with the rhythmic editing technique to create a sense of spatial progression through movement. The inclusion of walking shots propels viewers to visually move alongside the shots to travel across city scenes from narrow alleys to open waterfronts, leading to smooth spatial jumps and transitions. Short, rapid cuts render natural and seamless shifts between scenes, and elevate the energy and rhythm of city spaces during the festive season.

Such rhythmic editing unfurls the layered representation of spatial ambience. The fast-paced cuts function in sync with visual content, integrating elements such as crowds, festive decorations, and street performances into the visuals and constructing a city space with notable cultural recognizability. By regulating the rhythm of editing, the creator fluidly switches shots between physical locations, which develops non-abruptness but instead a “touring sensation of festive scenes” through rhythmic continuity, enhancing both narrative flow and emotional connection during spatial variations.

This rhythm-driven spatial linkage reflects the creator’s strength in coordinating the temporal-spatial relationship. The coherent editing rhythms naturally connect multiple signature locations associated with the holiday experience, crafting a city route centered on the joyous vibe. This narrative approach, where people’s movements drive the spatial transitions, turns the city into a shared celebratory environment that unfolds through the progression of editing rhythms.

In the short video “Night Tour in Chang’an: A Tribute to the Tang Dynasty and to Food”, narratives center around the theme of “night touring”, and editing rhythms are strategically adopted to thread Xi’an’s diverse nighttime scenes with cultural connotations. This approach presents a city image with a harmonious blend of modernity and traditions while highlighting its vibrant energy. The video opens at a

relatively fast pace, diving into scenes of lively night markets, food stalls, and crowds in motion (see Fig. 7). The quick shot transitions and upbeat narrative rhythm jointly create a dynamic, engaging visual experience.



Fig. 7

The creator pushes storytelling forward rapidly by placing such densely-packed editing rhythms across scenes, making viewers feel as if they were lost in the bustles and hustles of the Chang'an night market in person. Such editing-enabled precise pace control produces a compressed sense of time, delivering rich information within a short time frame, and amplifying the visual density of the night market. This demonstrates a highly perceptible city vitality through the language of the lens. In addition, the frequent scene changes facilitate building the so-called “viewpoint in motion”, where the viewer seems to follow each camera movement, touring quickly through the illuminated, crowded nightscape of Xi'an, which reinforces both the sensory connection and immersive experience of the city space.

Such fast-paced editing, as part of the rhythmic strategy, directs the viewer's attention to swiftly alternate between spatial scenes, including neighborhoods, festive locations, and activities intentionally placed by the creator in the short video. Viewers can follow on the single visual thread, moving across night market stalls, city squares, and commercial streets, and perceiving a rhythmic-driven jump-cut style of spatial narration. The high-frequency shot changes retain the overall visual tension while reinforcing the dispersed, multi-centered impression of Xi'an's nighttime layout. As a

result, the city takes on a visually fluid and structurally open image.

Upon depicting the modern streetscape in Xi'an through quick-paced editing rhythm, the video undergoes a noticeable slowdown in pacing by gradually switching the focus to historical landmarks, such as the Bell Tower, Drum Tower, and ancient city walls (see Fig. 8). Longer shots and more composed, placid visuals shape solemnity and tranquility. This rhythmic change takes the audience from viewing the invigorating modern city to perceiving the depths of its historical and cultural places. In all, these visual pauses recalibrate the viewing pace and divert attention from the vibrant street life to the city's tapestry of history, delivering a smooth emotional transition and a renewed spatial awareness.

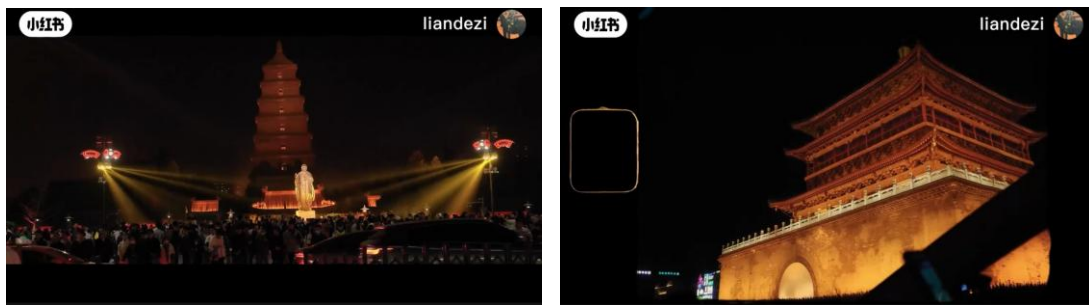


Fig. 8

This short video demonstrates how flexible control of editing rhythm, when supported by the precise arrangement of spatial scenes, enables a layered transition from urban bustles to the rich historical deposition, and from macro-level city views to everyday details. Such a rhythmic strategy helps the creator infuse varied city scenes into a coherent visual whole. As viewers follow the changing visual rhythm, they wander across Xi'an and come to collect more holistic insights into the city's multifaceted character.

The short video "Have You Ever Taken a Double-Decker Bus in Shanghai? This Is a City that Never Sleeps" applies the strategic design of "viewpoint in motion".

Commencing with public transport, the creator navigates viewers through the city space, shaping a textured spatial sensation in Shanghai spanning from daytime to nighttime, and from historical neighborhoods to modern architecture. Variations in editing rhythm emphasize contrasts and reflect the creator's conscious control over the interplay between city pulse and spatial structure.

The first half of the video unveils the placid atmosphere of vintage streets in Shanghai's old quarters, as driven by slow-paced editing and long takes. Viewers can sensorily get immersed in historical spaces, experiencing both cultural depth and everyday serenity (see Fig. 9) by following the creator's slow means of movement, such as walking and the ferry ride. In particular, the camera gently shifts between the city skyline and picturesque natural scenery during the sunset scene along the Huangpu River, articulating tender beauty during a moment of spatial and temporal transition. This pacing reinforces the relaxed tone of daytime Shanghai, driving viewers to forge an emotional connection with the city's time-honored spaces through stillness.

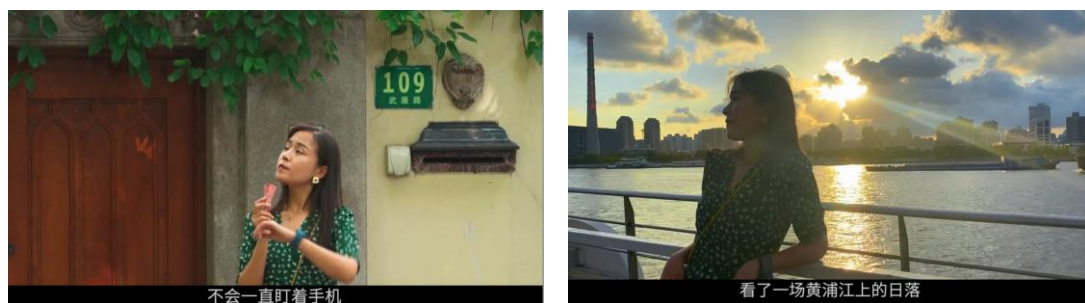


Fig. 9

This fast editing rhythm showcases Shanghai's modern essence and dynamism (see Fig. 10). Notably, during the double-decker bus ride, the shots switch rapidly to cover the bustling streets, skyscrapers, and urban landscapes along both sides of the Huangpu River. As the bus travels on the road, faster-paced shots depict Shanghai's heavy traffic and modern architecture, particularly the light-illuminated magnificent

skyline connecting the Bund and Pudong New Area. In this sense, the creator vividly conveys the motion and light effects of the city's nightscape, reinforcing Shanghai's image as a "sleepless city".

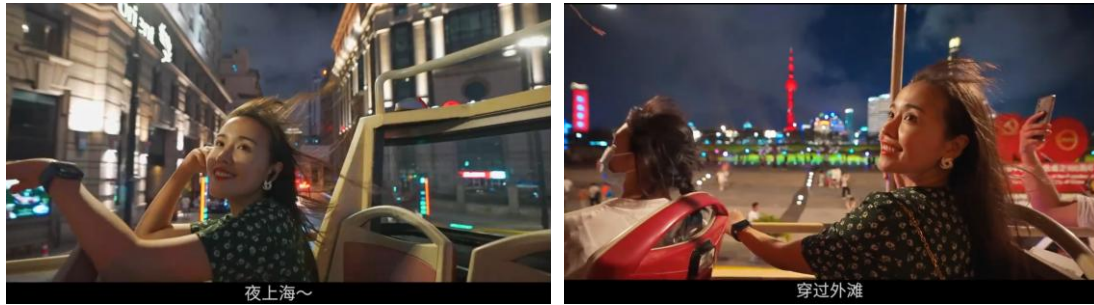


Fig. 10

Alternation between fast and slow editing rhythms aims to orchestrate and connect Shanghai's diverse city scenes in the video visuals. Rhythmic editing ensures clear, layered transitions between day and night, and between historical neighborhoods and contemporary buildings. This allows viewers to visually grasp the transformation and richness of the cityscape. Variations in editing rhythms between diverse scenes enhance the spatial flow and visual rhythm, facilitate the urban storytelling, and efficiently present the variety of Shanghai's spatial structure and lifestyle within a limited timeframe.

City cases illustrated in this study are distinguishable in their styles and preferences regarding the application of editing rhythm to scene transitions. In Beijing's case, the creator narrates at a consistently fast pace to link a variety of iconic Chinese New Year scenes, intensifying spatial continuity and cultural unity within the festal setting. Xi'an's video alternates between fast and slow rhythms to create layered contrasts between the vibrancy of night markets and the stillness of historical landmarks. In Shanghai's case, its editing rhythm assumes the function of orchestrating the temporal flows by adopting slower pacing during daytime and faster pacing during nighttime to reflect the city's transitions across time and space.

This manifests the distinct narrative priorities in each city's rhythm design: cohesive planning of festival atmosphere for Beijing; juxtaposition of historical heritage and modern lifestyles for Xi'an; and the interplay between urban rhythm and the sense of time for Shanghai.

Despite these varying approaches, all three cases represent the significance of editing rhythm in regulating spatial transitions by seamlessly connecting physical spaces, as powered by the creator's precise arrangements and rhythmic editing. Effective pacing adjustment and scene integration enable viewers to travel across multiple locations within a short viewing time, while maintaining a clear and structured visual logic underlying city image construction. This finding suggests that editing rhythm is a key pathway through which creators express a city's spatial rationale.

Shot combination in city image construction

Shot combination, as a significant film language commonly used by directors, cinematographers, and editors, is essential to constructing narratives, conveying emotions, and shaping the ambience (Rodgers, 2014). Individual shots, when carefully arranged and seamlessly interconnected, compose an exclusive viewing experience comparable to a fluid visual symphony that directs the viewer's gaze and emotional articulation (Figueiredo, 2015). Eyeing short video production, well-designed shot combinations allow creators to present the city's multi-layered character within rapidly switching visuals.

The power of shot combination stems from its multifaceted interpretation of shooting subjects through visual language. According to Wang (2025), shot combinations can skillfully unfold content in a layered, progressive way as per narrative demands. This textured arrangement reflects a well-planned creative strategy in the short video

context. Creators can scrupulously design shot sequences to orderly seize varying spatial dimensions, initiating stepwise city explorations for viewers and progressively enriching and deepening perceptions of a city's overall character.

Shot combination is able to communicate emotions with subtlety, which stands for its core function in boosting emotional connections. The considered design of visual language and content arrangement in short videos can relate the video content to the viewer's emotional experience (Zhan, 2024). This enables viewers' progressive immersion in the creator-enabled emotional narrative. Such skillful shot combinations transform the city from a "concrete jungle" into an affective lived-in space, engaging viewers in more constructive, in-depth emotional experience through short video production.

This section dives into how shot combination bolsters short-video-based city image construction in the three selected cities, and examines the characteristics and effects of shot combinations in each city case through a comparative analysis.

Shot combination and multi-layered city display

Short video creators visually present a city's variety and naturally integrate its multidimensionality by orderly coordinating and shifting shots where appropriate. A city image is, by definition, a complex construct encompassing the physical environment, historical traces, cultural symbols, and modern evolutions. Shot combination instills viewers with a city's multi-layered information within a brief viewing time of short videos, thereby cultivating viewers' visual synthesis of the city ethos.

In the short video "Roadside Observers", Beijing's multifaceted image is presented by combining various shots. The creator condenses a rich variety of shot transitions in

just one-and-a-half minutes, silkily bridging Beijing's historical landmarks, modern landscapes, cuisine culture, and entertainment activities (see Fig. 11). Such a shot combination, as a blend of temporal and spatial shifts, leads the audience to acquire insights into the city's culture and lifestyle.

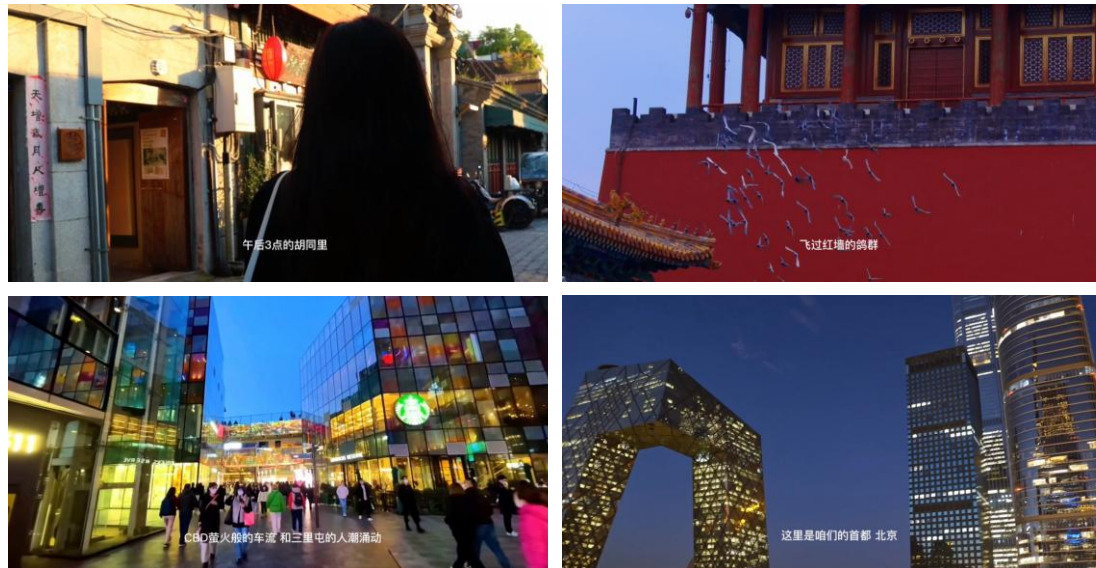


Fig. 11

The micro-level shot combination in food scenes proves particularly effective in the short video in which storefronts of traditional Beijing snacks, such as long-established restaurants and street stalls, are presented alongside dining scenes and close-up shots inside the eateries. These combined shots grasp the richness and diversity of Beijing's culinary offerings, visually featuring Beijing dishes such as Jianbing guozi (a type of Chinese breakfast crepes) and Zha Jiang Mian (Beijing Fried Sauce Noodles) in a way that highlights their appeal. This thus enables viewers to experience Beijing's culinary culture from multiple perspectives, investing the city image with a warm and down-to-earth atmosphere.

Additionally, macro-level shots are leveraged to capture the peacefulness of Beijing's stand-out hutong culture. A key segment from the video features elderly men strolling through hutongs and interacting with animals such as koi fish and sparrows, standing

as a vivid display of Beijing's distinctive community culture and everyday tranquility. The shot combination, portraying enriched sensory experiences from afternoon leisure in hutongs to the nighttime entertainment, including stand-up comedy, or performances in theaters and concerts, allows the audience to feel the rhythm and flow of the city from different perspectives.

As discussed above, the creator intertwines scenes of hutongs, modern landscapes, and traditional cultural activities, displaying the fusion of history and modernity in Beijing through macro-level shots. Viewers can deeply sense the multi-layered culture and daily pulses of the city with micro-level shots. Overall, such a shot combination effectively transcends the physical barriers of conventional urban representation, facilitating the audience to appreciate Beijing's unique charm in varying dimensions.

Xi'an, as one of China's ancient capitals and a pivotal cradle of traditional culture, is home to a wealth of historical and cultural elements and modern progress. In a short video about Xi'an, "Enduring Love in Chang'an", the creator skillfully pursues the shot combination to blend its ancient and contemporary elements, going beyond the city's temporal and spatial boundaries to showcase its multi-layered image.

In this short video, the creator bridges antiquity and modernity through seamless shot transitions (see Fig. 12). A typical shot combination appears at the onset of the video, where the creator, wearing modern attire, steps into a hotel lobby well-equipped with modern facilities, embodying Xi'an's contemporary metropolis identity. In the next scene, the creator, dressed in exquisite traditional clothing, immediately transitions to a classically-styled hall, through a simple turn of a corner. This shot transition uses a time-space leap to highlight Xi'an's intermingling of ancient culture and modern civilization.



Fig. 12

This technique of challenging temporal boundaries through the shot combination considerably helps present Xi'an's city image. The city embraces the co-occurrence of historical depth and flourishing modernity. From alternating shots of modern architecture and ancient relics to the creator's costume changes, audiences are guided to traverse the millennia-old history in Xi'an within a short span of viewing. This design underscores the city's vitality and openness in its modern progression.

The video's depiction of classical elements further strengthens the city's cultural image. Shots of iconic ancient buildings, such as Xi'an City Walls and the Giant Wild Goose Pagoda, are covered in the video to allow for the audience's perception of Xi'an's long-established historical legacy as an ancient capital. Integrating these classical elements with modern hotels and commercial districts weaves the cultured charm of Xi'an's classical civilization with its modern renewal.

Xi'an's multi-layered representation is enabled by masterful transitions of shot combination, too. The intertwining of history and modernity is reflected both in the selection of classic architecture and scenes and in the narrative strategy of shot combination, which transcends the limitations of time and space. This enables the audience to perceive and appreciate Xi'an's overall image in further depth in a short period.

Short video creations embody the multidimensionality of Shanghai, a microcosm of

China's modernization and internationalization. In the short video "Ten Years On, I Still Miss the Shanghai Depicted in *Tiny Times*", the creator combines film scenes with live-action shots, introducing more literary and sentimental layers to Shanghai's city image.

In this short video, the creator applies shot combination to blend the depiction of Shanghai in the film *Tiny Times* with the real-life urban landscapes, creating a mix of fiction and reality (see Fig. 13). Locations typical to the film, such as the Bund and the Oriental Pearl Tower, as well as real-life footage of Shanghai's bustling and international identity are all incorporated into this short video. This shot combination works brilliantly by mingling virtual and real-world scenes and thus propelling Shanghai's multi-layered image to emerge.



Fig. 13

The film scenes set up in Shanghai reflect its prosperity, fashion, and modernity, while the creator's live-action shots extend these fictional settings into reality, enhancing the audience's emotional resonance with the city. For instance, the video shots of the Oriental Pearl Tower echo the film scenes, which reinforces Shanghai's city identity

as a symbol of international urbanism. Additionally, some shots are extracted from locations in the everyday life of Shanghai, showing that the city is both a place lined with skyscrapers and a warm, lively urban space teeming with everyday human experiences.

Furthermore, the shot combination adds literary qualities and emotional depth to Shanghai. Multiple city scenes are skillfully interconnected with the film as a narrative thread, turning Shanghai's image into a continuation of the film storyline while imbuing it with deeper cultural connotations. This strategy enables viewers to go beyond perceiving Shanghai's physical architecture and cityscape towards developing emotional and cultural insight into its unique charms.

The fusion of film and real-life footage fully displays the multi-faceted essence of Shanghai for viewers' layered perception and interpretation. The shot combination effectively conveys Shanghai's spatial breadth and cultural depth as a global city in the short video.

Creators recognize that shot combination is crucial to presenting the city from multiple perspectives, as seen from all the short video cases about the three cities. Creators break away from static visual representation using horizontal changes in lens distance, vertical shifts in time and space, and alternations between interior and exterior scenes.

These shot-structured visual strategies all lend spatial depth and temporal continuity to city image representation, allowing the city to be both perceived as a physical entity and an intricate vessel of social, cultural, and emotional experiences. Shot combination makes it possible to grab viewers' visual attention and emotional engagement within a short time. In practice, it leads viewers into the city's multilayered image and enables the simultaneous transmission of spatial information

and cultural meaning via short video creations.

Meanwhile, the approach to shot combination in each video case reflects the target city's exclusive urban character and cultural context. Beijing's video case features a sequence-based structure to portray a city juxtaposed by everyday rhythm and cultured traditions to emphasize spatial movement. Xi'an's video case prioritizes temporal shifts and stylistic contrasts, filming shots that switch freely between the modern and historical landscapes to shape a visually striking narrative. Shanghai's video case innovatively blends cinematic and real-life imagery to layer its cultural depth and emotional richness, constructing a globally modern and aesthetically complex imagery. These differences reflect the creators' adept ability in content curation and orchestration of visual elements in alignment with each city's identity and narrative intention.

Shot combination and emotional connection with the city

In short video production, a city is not just defined as a physical entity made up of geographical space and architecture, but also as a compelling emotional field. Thoughtfully composed shots taken by creators can lead to a layered emotional connection between viewers and the city, and ultimately cultural identification and resonance. According to Block (2020), the shot combination strategy can forge an emotional progression that binds the audience to the short video upon mindful arrangements of specific city scenes and everyday moments, transcending the common representation of physical space to unravel the emotional presence deep within the city.

When shaping Beijing's city image, the blogger "Lili Shanhe" demonstrates her emotional tie with the city by shooting the short video "You Have to Climb the Great Wall Once, Feeding Cats, Eating Popsicles, and Admiring the Vast Scenery". The shot

combination featured within the video documents her ascent to the Great Wall, and more crucially, intertwines her physical challenge with an emotional outlet by revealing her emotional navigation and deploying skillful shot transitions. This forges a deep relatability between the individual and specific settings of Beijing.

The video opens with multiple angles showcasing the creator's challenging climb of the Great Wall (see Fig. 14). Upward shots emphasize steep inclines, enabling viewers to intuitively sense the ascending difficulty, while side shots capture the creator's body movements, further underlining the physical exhaustion. Aerial shots present the sheer terrain to heighten the visual impact of the Great Wall's ruggedness. This multi-lens approach plunges viewers into this experience fraught with challenges. The combined shots embody the emotional buildup, such as mental ups and downs, resulting from this demanding experience, in the viewing experience.



Fig. 14

When the creator finally summits the Great Wall, the shot combination changes. Specifically, as the creator points into the distance, the lens shifts to showcase the natural grandeur outside Beijing, as outlined by undulating mountains and the meandering Great Wall. This presents a shift in the visual landscape while conveying the emotional shift from fatigue and struggle during the climb to relaxation and pride upon reaching the summit. The shot combination technique facilitates this shift, and the creator's personal experience resonates with Beijing's expansive scenery, marking this emotional journey from difficulty to spiritual relief.

These shot choices and arrangement all transform Beijing from a city profound in historical heritage into an all-encompassing space for emotional experience. The Great Wall, as this national capital's emblematic representation, bridges emotional resonance between the short video creator and viewers (Shiping, 2025). Following the shifts in shots and gradual build-up of emotions, viewers are drawn into the creator's emotional world, during which the city image emerges vivid and inspiring through this affective connection.

In the short video “Enduring Love in Chang’an”, Xi'an's city image is vividly portrayed through the dynamic accumulation of emotions. The creator masterfully shapes an emotional atmosphere through shot combination, effectively engaging the audience to feel Xi'an's liveliness and vibrancy through visual language (see Fig. 15) in a limited viewing journey.



Fig. 15

As shown in this short video, the creator is participating in art events during Xi'an's nightlife, whose rapid shot transitions capture the colorful dynamism across the city from multiple angles. Various scenes, such as street performances and cultural exhibitions, are connected via the shot combination to convey the city's sustained vitality and passion. Despite the absence of explicit depictions for historical and cultural landmarks, the dynamic integration of shots immerses viewers in the distinctiveness of its urban atmosphere and rich cultural legacy.

The progressive emotional articulation by the creator leads the audience into forging affective ties with the city. As night falls and nightscape activities unfold, the creator accelerates shot transitions, resulting in the corresponding shift in the audience's emotional perception. A wealth of emotions is ultimately built up by repeated transitions in shots and scenes. This unique shot arrangement also enables viewers to sense the creator's growing emotional relatability with this fascinating city.

Emotional articulation is more focused on shaping a collective atmosphere, as observed from the shots recording Xi'an's nightlife. A combination of street scenes, night markets, and art events fully displays the emotional warmth of Xi'an. The creator conveys the city's bustle and vitality through visual language, recreating its exclusive allure glowing under the night sky. This shot combination strategy synchronizes with the city's emotional and visual imagery, where emotions and scenery complement each other, presenting a culturally rich, resonant image to the audience.

Shanghai, as a modern metropolis, is also invested with multi-layered emotional depth in the short video "What Kind of City Do You Think Shanghai Is?". The creator employs emotional storytelling and shot combination to portray the complex emotional layers within the city, and tightly interweaves the city with personal emotions through extensive shot transitions. This cultivates the natural immersion of viewers in the emotionally-layered viewing experience of this city.

The video begins with the creator walking through the busy, neon-lit streets, revealing Shanghai's urban landscape to the audience in glimpses. This scene, captured through fast-paced shot transitions, conveys the bustle of the metropolitan city. High-rises, street pedestrians, and iconic labels like Lujiazui symbolize Shanghai's prosperity and dynamism as a global metropolis. The creator illustrates the individual's triviality and

solitude in this immense city rife with survival strain and pragmatism (see Fig. 16) with a meticulously planned shot combination.

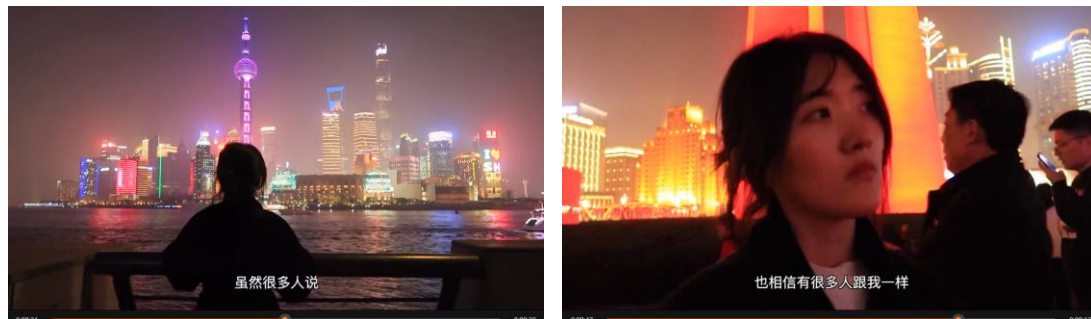


Fig. 16

However, as the shots delve deeper into the creator's inner thinking, she presents notes of hope through further shot transitions. In the Lujiazui-based scenes where the creator shares her individual experience, Shanghai is no longer portrayed as a representation of the indifferent urban landscape, but as a vehicle of in-depth human touch. As she navigates through the vibrant streets, the camera captures interactions of passers-by with the city, and deepens this emotional bond through detailed framing. The intersection of Lujiazui's iconic architecture and the creator's individual emotions endows these frigid city structures with new meanings, symbolizing aspirations and potential for the future.

By the end of the video, the creator's scrupulous observation of Shanghai emerges as the core in visual expression, communicating the emotions deep within it. The creator imbues the city's everyday life with distinctive emotions through the arrangement and transition of shots. As the creator's individual feelings are adeptly interwoven into the city image presented by shot combination, viewers get to perceive the city's inclusiveness and complexity in an emotional dimension, and naturally develop resonance with the city.

In this way, Shanghai's city image is constructed as a richly-layered entity abundant in

emotions and humanistic warmth rather than an emotionally distant collection of architectural symbols. The shot combination presents the city's flourishing and bustling ambience, while the nuanced emotional expression allows viewers to sense the warmth and expectations within it, and thus uncover that Shanghai's ethos is vibrantly and profoundly conveyed by such an emotion-space connection.

Although short video creators across the three cities vary in their strategies of shot combination, visual coordination and arrangement are all applied to transform the city space into a dynamic field of emotional expression, establishing profound relationships between viewers and the city.

In Beijing's video case, an individual's emotional journey during the Great Wall climb weaves together personal feelings and collective memory. In Xi'an's video case, a variety of activity-based shots jointly shape the shared emotional atmosphere, lighting up the night with warmth and humanistic presence. In Shanghai's video case, scenes associated with solitude and hope are integrated into an emotional whole, cultivating inclusiveness, variety, and tenderness in the city image. While the video cases differ in emotional dimensions, the shot combination assumes the role of evoking viewers' emotions and reinforcing the city image in all of them.

In terms of specific strategies, Beijing's video case tends to arouse emotional tension through spatial challenges, using the camera lens to reflect personal physical experience and highlight emotional bonds between individuals and the city. Xi'an's video case focuses on rhythm-driven emotional buildup, where the lens serves to build interpersonal relationships and forge ties between people and diverse activities, thereby constructing an immersive city atmosphere. As for Shanghai's video case, it frequently shifts between cityscapes and individual experiences, turning the camera shots into carriers of subtle emotions and adding a tender human touch to an otherwise indifferent urban image. In this sense, shot combination emerges as both a

visual expression tool and emotional core in the creator's strategy, which helps with shaping mood, organizing content, and communicating the city's ethos.

Visual special effects in city image construction

As powered by evolving digital technology, visual special effects have found their extensive applications across various visual media, including short video platforms. Visual special effects allow creators to break free from realistic constraints, presenting city images in more diverse ways and supporting audiences with a fresh perceptual perspective. As elucidated by McClean (2007), visual special effects grant creators with higher expressive freedom by enabling reality-based remakes and imagination, thereby diversifying the possibilities of visual representation.

Visual special effects in short videos can help shape a city's distinctive visual style and be tailor-made to fulfill the intended outcome, such as highlighting the director's strategic intent (Hopkins, 2008). Modern cities may employ technology-driven visual special effects to convey innovation and vitality, while historically rich cities tend to express cultural depth with conventional visual special effects. The two cases both facilitate viewers to forge an emotional bond with the city and sense an increased visual impact under the effects of visual enhancement.

Visual special effects are also fundamental to conveying and supplementing information. Creators often face a challenge in giving full play to the historical, cultural, and economic dimensions of a city in short videos that feature a limited length. Visual special effects offer a variety of means of expression in answer to this challenge (Figueiredo, 2015). Creator, by furnishing key information through visual special effects, navigates viewers to probe into the cultural meaning underlying the city image.

The following section analyzes how visual special effects are applied in the three cities' video cases for imagery enhancement and information supplement, and their respective contributions to city image construction.

Visual special effects in city imagery enhancement

A city's distinct visual symbols, cultural representations, and spatial structures can be further amplified or reinforced through visual special effects that can significantly elevate the uniqueness and recognizability of city images, helping viewers recognize and keep in mind the city image within a short timeframe. This makes the city no longer confined to physical spaces or real-world architecture. Instead, the city image takes on greater expressive flexibility and imaginative potential thanks to the creator's visual special effect design (Wang, 2024). By recreating or extending real elements, the city image in short videos transcends everyday experience and emerges as a more visually impactful and sensorily appealing symbolic representation.

As for Beijing's city image, visual special effects are utilized to enhance the visual impact, and aesthetically infuse the city's iconic architecture with musical performance gestures, thus renewing the city image (see Fig. 17). In the short video "In Tune with the World Here", the creator ingeniously renders green screen keying techniques and visuals effects to seamlessly blend Beijing's landmark buildings with the ongoing musical performances. This augments the city's visual symbols, thereby offering a new perceptual approach to viewing the city, in which the interplay between music and architecture reinterprets its cultural and artistic depth.



Fig. 17

In the video, the Erhu playing gestures are creatively combined with the White Pagoda from Beijing, creating a stunning visual imagery where the pagoda itself seems to be a component of the Erhu instrument. Similarly, directing movements of a piano performance is combined with the eaves of a palace, as if the architecture were part of the performance. The creator does not merely showcase the buildings but invests them with dynamic expressiveness by applying these visual special effects, thereby breathing life into the long-established architecture. This use of visual special effects transcends the static nature of landmarks by transforming them into an integral part of the city's culture and art.

Visual special effects upgrade Beijing's landmark structures into strong visual symbols, boosting the construction of its cultural character and artistic temperament, and blending its historical and cultural heritage with its modern artistry. In this way, viewers can experience the city's multilayered and rich connotations within the short viewing span.

Xi'an, as a well-known historic and cultural city, has its natural spectacles and cultural variety further accentuated by visual special effects in short videos. In the short video "Don't Doubt It, This is Xi'an", visual special effects are artfully employed to produce a fog effect. Detailedly, followers can feel the morning splendor in Xi'an, especially the tranquility and beauty of the city's outskirts (see Fig. 18), through a misty, dreamlike scene.



Fig. 18

The fog effect is pivotal to shaping a surreal, fog-filled atmosphere where viewers feel as if they were inside a picturesque painting. The moment when a flock of birds is flying across the lake is particularly striking, creating a feel that their flight path is meticulously adjusted upon the adoption of visual special effects to ensure a seamless integration with the overall scene. This brings forth greater visual harmony, which in turn enhances the viewer's experience.

Additionally, visual special effects are applied to adjust the color tones in visuals, imbuing the entire scene with a cool-toned aesthetic and amplifying the misty and tranquil atmosphere of the early morning. In this way, the creator evokes a classical ambience that aligns with Xi'an's historical and cultural identity. Thus, Xi'an's spectacular natural landscape is sublimed into a lyrical and dynamic visual experience marked by the mist and flying birds in the video. Such visual enhancement conveys Xi'an's cultural richness and harmonious relationship with nature, and constructs a more ethereal and aloof city imagery.

As for the video case of Shanghai, this modern international metropolis, visual special effect strategies often stress the integration of technology and a futuristic sense, as is showcased in the short video "A Giant Moving 'Captain Shu' Appears on the Streets of Shanghai". A variety of visual effects are used in this video to introduce Xiaohongshu's cartoon mascot "Captain Shu" into Shanghai's urban space, blending

the animated character with real-life city buildings to fuse reality with virtuality (see Fig. 19).

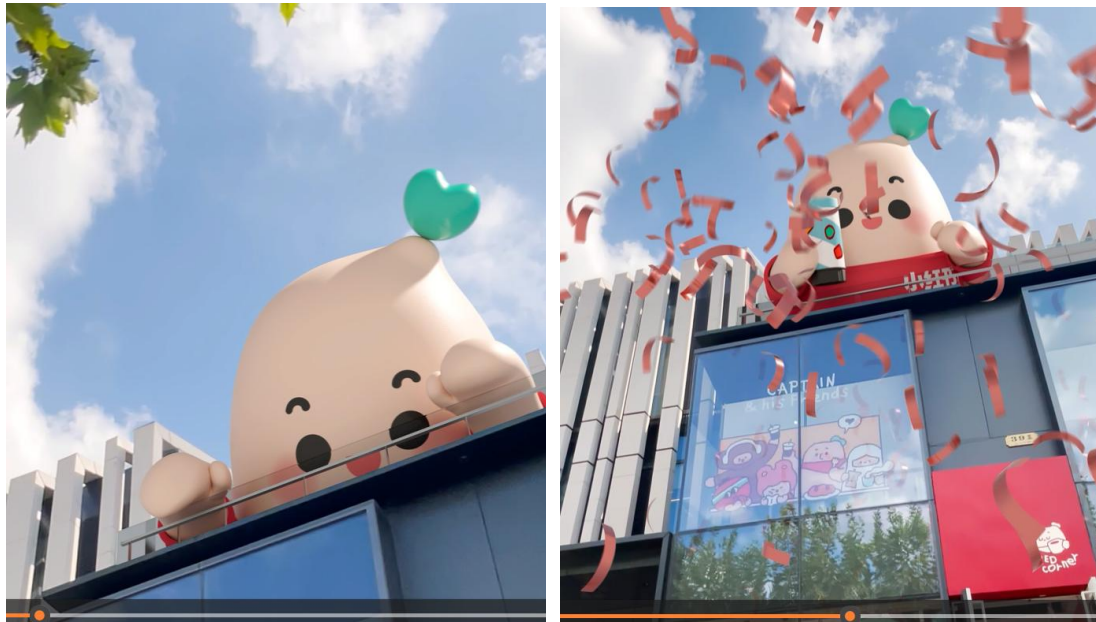


Fig. 19

In the video, “Captain Shu” appears to “crawl” out from behind a building due to the ingenious use of visual special effects, which organically integrates Shanghai with new media culture. Its fusion with the city ambience reflects Shanghai’s multicultural background as a modern metropolis and a global hub for culture-technology integration, which leads the audience to perceive an atmosphere of fun and accessibility during viewing.

Notably, rich visual special effects deliver a more lively and interactive urban space on the visual level during a celebration scene of the “color-throwing festival”. This visual enhancement technique allows viewers to feel Shanghai’s modern vibrancy while conveying a relaxed, entertaining city image by introducing this cartoon mascot that is brought to life by dynamic visual effects. This perfectly aligns with Shanghai’s image as a city abundant in innovation, fashion, and energy.

The seamless blend of virtual and real elements enabled by visual special effects challenges conventional representations of urban landscapes, offering viewers a fresh perspective to experience the city's modern features. Shanghai's city image is thus no longer confined to skyscrapers and bustling urban life. On the contrary, it comes across as multi-dimensional and futuristic through the integration of innovation and entertainment elements.

This study identifies common strategies and city-specific differences in visual enhancement by examining the use of visual special effects in short videos about Beijing, Xi'an, and Shanghai. The commonality lies in the purpose of transcending the limitations of real-world imagery. Visual special effects, by boosting more concrete and artistic expressions, amplify distinctive urban features, improve the recognizability and memorability of each city's image, and serve as visual tools to activate cultural meanings, emotional atmosphere, and narrative structure.

Each city's case boosts exclusive techniques and expressive orientations. In Beijing's video case, iconic architecture is intertwined with performative actions, using aesthetic transformation to strengthen the interplay between buildings and cultural expression. This highlights the city's textured cultural heritage and artistic sensibility. Xi'an's video case focuses on the creation of natural ambience and the extension of historical mood, producing a hazy, tranquil visual experience that immerses viewers in a poetic perception of the city with visual special effects. Shanghai's video case adopts a more technology- and entertainment-driven approach to convey its modernity and energy, as shown from its cartoon imagery and dynamic visual special effects. This effectively constructs a city image with tech-inspired international appeal.

Therefore, how visual special effects are utilized to reinforce city imagery reflects diverse creative paths shaped by varying city identities, cultural characteristics, and creators' strategies. Overall, this representational diversity suggests more possibilities

and creative potential in city image construction.

Visual special effects and visual supplementation of city information

The condensed timeframe of short videos often makes it difficult for creators to bring a city's history, culture, economy, and other dimensions into full play. Visual tools can alleviate this issue by using diverse forms of presentation and supplementing content.

In terms of Beijing's video case, special effects are significant in conveying information, as they help viewers perceive the city's development and transformation in a short timeframe. In the short video "AI My Capital: 12 Data Points on Ecological Progress", the creator applies dynamic subtitles and icon labels to graphically present Beijing's ecological improvement and environmental protection progress over recent years, especially through the dynamic presentation of data.

Visual special effects are used to display variations in data related to pollution and waste management (see Fig. 20) from recent years in the video. Viewers can clearly identify trends such as reduction of pollutants, increase in species diversity, and water quality improvement through animated icons. These data are closely integrated with real-world visuals of Beijing. For example, when demonstrating the enhanced species diversity, footage of Beijing's ecological environment is set as the background. This turns the data from abstract numbers into integral parts of viewers' visual perception of the city. In this way, visual special effects help bridge visual imagery and urban development data, allowing viewers to internalize such information into their comprehension of Beijing's city image with more ease.

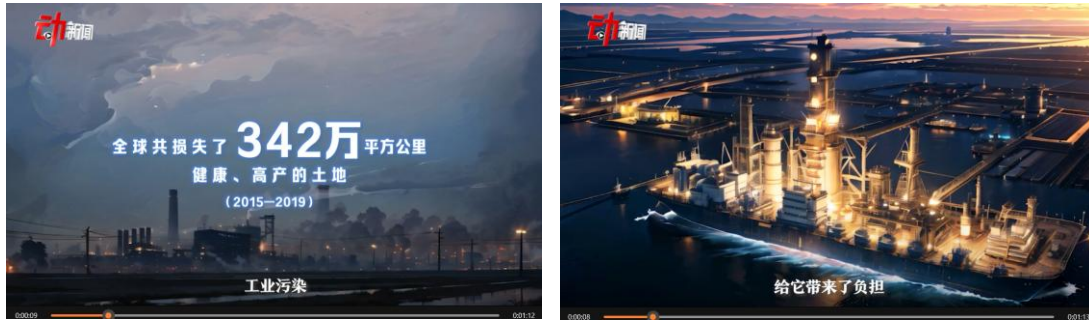


Fig. 20

Additionally, dynamic subtitles provide substantial support to information transmission in the video by carrying data and applying animation and stylistic adjustments to lend an upbeat rhythm to the visuals, helping viewers keep pace with the video content. The interaction between subtitles and icons furthers viewers' knowledge of the city's ecological progress and delivers smoother, layered pacing in the video.

Visual special effects help articulate a more comprehensive picture of Beijing's city information as its visual supplements, through which creators enable viewers to perceive Beijing's ecological improvements. As a result, they are more assured of the city's sustainable development capacity. Such effective information conveyance ensures that Beijing's city image moves beyond the display of mere visual landscapes towards substantive representation of environmental and societal progress.

In Xi'an's city image construction, visual special effects are creatively used to supplement information, particularly in the short video "If You Want to Write About Xi'an, You Can't Just Write About Xi'an". This video features the seamless fusion of subtitles and visuals, which recreate a distinct vintage. The poetic expression of visual special effect subtitles presents Xi'an's historical and cultural knowledge and builds a picturesque ambience, thereby constructing a multidimensional and emotionally resonant (see Fig. 21) city image.

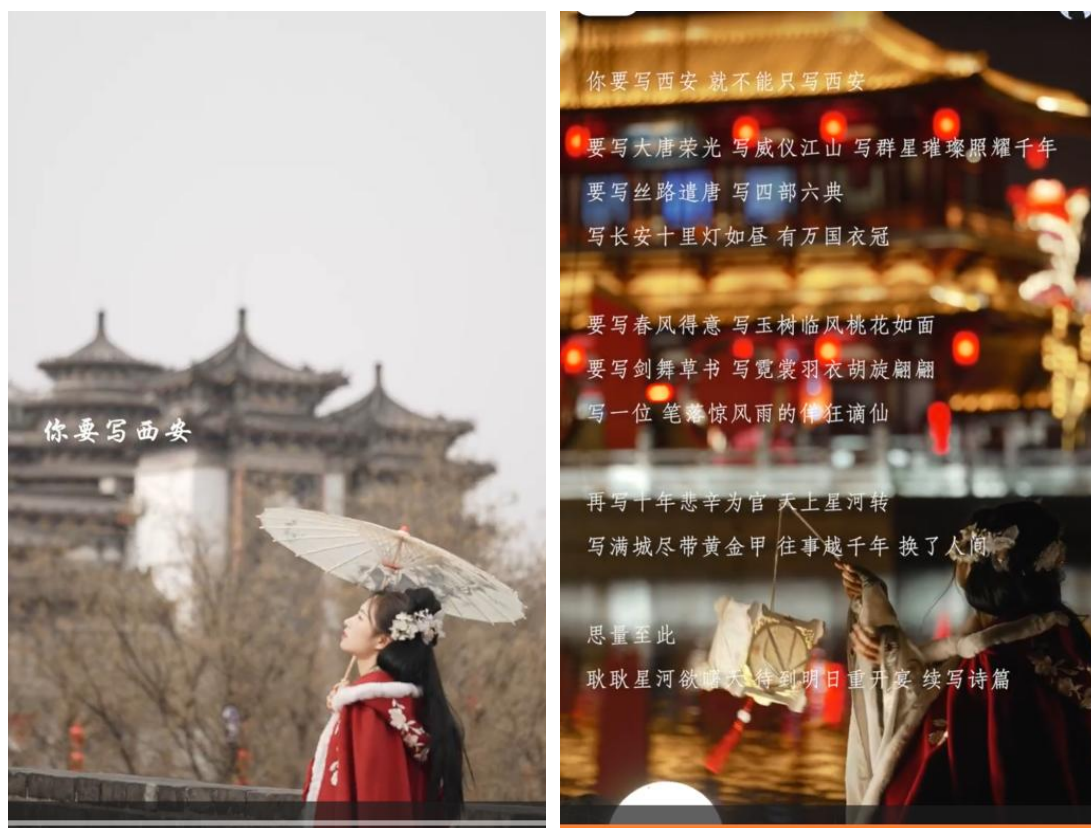


Fig. 21

The special effect subtitles in the video are blended with the ancient cityscape of Xi'an, dually illustrating the city's scenery and forging a profound interplay between the city and the visuals. The antique style in subtitles aligns with Xi'an's cultural context as an ancient capital, thereby enhancing the cultured and historic vibes. The poetic content in subtitles portrays the city image with metaphors and depictions of Xi'an's history. Consequently, the audience can appreciate the city's spiritual essence through the visual and textual impact. Xi'an's city image transcends mere architectural and landscape displays, and the power of texts augments the transmission of information. Viewers, by visually supplementing information, gain insights into the city's exclusive cultural lineage and historical heritage.

Producing short videos about the modern metropolis of Shanghai encounters the hurdle of manifesting densely packed information about its multi-faceted urban landscape. In the short video "A More Stunning Landmark Than the Bund, Yet

Overlooked by Tourists”, the creator guides viewers’ attention to key elements in the frame by supplementing historical and background information, such as building annotations. In this way, viewers attain a lucid comprehension of the architectural culture in Shanghai.

A cluster of buildings steeped in prolonged history along the Bund functions as the video focal point. However, the large quantity of them may make it difficult for viewers to grasp the specific features and historical context of each building from the footage alone. Thus, the creator has chosen to apply special effect annotations containing crucial information, such as building names and construction years, directly to the visuals. For example, as the camera lens sweeps across the former Shanghai Tram Company building, the special annotation intuitively indicates that it was built in 1908, enabling viewers to instantly obtain the related historical background information (see Fig. 22).



Fig. 22

These special effect annotations draw viewers’ attention and simplify information delivery. In short videos without special effects, viewers often identify the difficulty

in keeping their attention on key content within the complex urban visuals. By adding visual special effects in forms like annotation, creators enable viewers to concentrate on architectural information with higher significance, and ensure a more explicit representation of Shanghai's historical and cultural information. This helps lower the fragmentation in short video content and facilitates viewers to develop a better knowledge of Shanghai's unique urban characteristics, such as the integration of history and modernity.

As a result, Shanghai's architectural culture is more effectively conveyed, resulting in the enrichment and deepening of the city image, and an improved viewing experience thanks to the visual special effects that supplement critical information.

Short videos of Beijing, Xi'an, and Shanghai keep their overall strategy consistent with each other. Creators from these three cities translate intricate city information, including history, culture, and spatial layout, into easy-to-read visual information through visual special effect formats such as subtitles, charts, and annotations within a limited time frame. This approach both ensures efficient information efficiency and compensates for the inherent spatial constraints in short-video narratives, thus imparting city images with greater knowledge density and cultural depth. Visual special effects, whether presented through dynamic subtitles that strengthen temporal and spatial coordinates, or graphic elements that embody data and landmark information, emerge as key vessels that communicate city information to the audience.

However, the means of implementation and visual expression differ. Beijing's video case emphasizes the integration of visual special effects with its urban governance image, using data charts and logically organized dynamic annotations to highlight its modern management progress. Xi'an focuses on cultural ambience, introducing visual special effects into traditional imagery and poetic expression to endow information

with historical weight and human warmth. Shanghai adopts a more utilitarian style, using guide-like subtitles and map annotations to emphasize the fast pace and clear functionality of a modern metropolis. These differences not only illustrate the adaptability of such techniques to different contexts but also reflect creators' strategic awareness and aesthetic preferences when constructing a city's informational image. With the help of such differentiated applications of visual special effects, each city develops its own distinctive visual narrative path while supplementing pivotal information.

Conclusion

Post-production techniques are central to city image construction in the short video context. This chapter takes a thorough analysis of short video case studies in the three representative cities, Beijing, Xi'an, and Shanghai, to clarify how editing rhythm, shot combination, and visual special effects, as the three core elements of post-production, effectively shape and convey city images.

Editing rhythm, with its variations between different paces, reveals the multifaceted nature of each city. Beijing's video cases combine faster with slower editing rhythms to showcase the fusion of antiquity and modernity. Xi'an's video cases visualize the balance between historical heritage and modern features. Shanghai's video cases mainly apply fast-paced editing to highlight its modern prosperity and dynamism as an international metropolitan city.

Shot combination is critical to both multi-layered presentations and emotional connections, as displayed in the short videos of each city. Beijing's video cases emphasize the intertwining of modern allure and traditions. Xi'an's video cases display the alternation of different periods. Shanghai's video cases apply this strategy to represent its modernization and internationalism. In terms of emotional connection,

each city's video cases differ in exhibited characteristics: personal experiences in Beijing; collective vibrant atmosphere in Xi'an; and urban interactions in Shanghai.

The use of visual special effects further diversifies the representation of city images in each case: highlighting cultural and artistic representation in video cases about Beijing; stressing the construction of a historical ambiance and showcasing natural beauty in Xi'an's video cases; and underscoring technological sense and futurism in Shanghai's video cases. Visual special effects also propose valuable solutions to information supplementation in each city case, ranging from data visualization for Beijing, to cultural information integration for Xi'an, and historical building annotation for Shanghai.

To conclude, a wide variety of post-production techniques studied in this chapter indicates creative potential for shaping city images in varying ways, showcasing each city's stand-out charm and cultural essence in short videos.

Chapter 8 Conclusion

This study has explored how techniques such as color, light and shadow, sound, characters, and post-production are applied in the construction of city images. Taking Beijing, Xi'an, and Shanghai as case studies, it focuses on the strategies used to construct city images in short videos on the Xiaohongshu platform. Moreover, it conducts a comparative analysis of these technical approaches in short video cases to reveal how creators construct city images under the operational logic of short video platforms.

The following sections first summarize the research process and main findings, and then outline the theoretical, methodological, and practical contributions of the study. It also reflects on the research limitations and suggests possible directions for future research, aiming to deepen the understanding of how city images are constructed in the era of short videos.

Key findings of the research

This study, centered on the core question of how creators construct city images on short video platforms, comparatively analyzes the image construction strategies for the three cities in the digital media context. Specifically, it puts forward a multidimensional analytical framework threaded by five types of representational techniques, taking Xiaohongshu as the research platform and conducting case studies of Beijing, Xi'an, and Shanghai. It also delves into short video content through qualitative content analysis to uncover the interactive mechanisms among city resources, creators' strategies, and media expression during city image construction.

The overall research proceeds from perceptual foundations to strategic approaches and finally to technical implementation. This study dually stresses the detailed

construction of city images and examines the differences in image representation and construction through creators' portrayals of cultural elements from each city. It assesses the similarities and differences in city image construction through a comparative approach to various cities and content types, revealing how these city images are represented under the distinctive communication rationale of short videos, and identifying four main findings as follows.

First, varied strategies are applied to construct images of different cities in short video production. They are deeply influenced by each city's historical background, cultural resources, and communication positioning. For example, short videos about Beijing often center around constructing its image as a political hub and cultural capital, emphasizing its solemnity and historical profundity. Xi'an's video cases often highlight local traditions, intangible cultural heritage, and everyday street life, presenting an accessible, regionally rooted character of the city. Video cases depicting Shanghai revolve around themes of fashion, urban lifestyle, and personal emotions, fusing dynamism with modernity in its image construction.

Although strategic priorities differ in each city's video cases, they all embody the same media logic and construction mechanism of short videos, which suggests that city image construction is essentially grounded in a shared framework of digital communication, embodying differences within commonality.

Second, applications of representational techniques, such as color, light and shadow, and sound, substantially differ in the three cities' short videos in line with creators' expressive preferences. These preferences unveil creators' subjective judgments in aesthetic focus, content planning, and media adaptability. For instance, some creators prefer shaping a warm ambience with softer tones and natural lighting, while others accentuate strong color contrasts and visual tension to capture viewers' attention. These non-random differences stem from individual interpretations and stylized

reconstruction of city characteristics. This highlights that creators, besides producing content, actively engage in the construction of city images, and their choices significantly shape viewers' perception of cities.

Third, short videos shift a city's role from a conventional backdrop or geographical space to an object implicitly perceived and experienced by viewers, and from an abstract aggregation of places to a confluence of specific settings and emotions. Viewers can feel the rich atmosphere and rhythm of city life through creators' visual language, content design, and precise control of pacing. In this way, a city image presented in media forms extends beyond a visual symbolism towards a vessel for emotional projection and cultural identification, as commonly observed from video cases for the three cities. It reflects a broader trend in which short video creations are reshaping people's perception and engagement with city places as well.

Fourth, city images are not statically produced, but dynamically constructed by the synergy of creators, media platforms and technologies, and urban realities. From one perspective, a city's level of development and available cultural resources influence how it is presented in visual content. From another perspective, advancements in media technology continuously enrich the expressive possibilities for creators. Tools such as filters, editing rhythms, background music, and transitions diversify and elevate the appeal of city images. This dynamic process, driven by various forces, is widely seen from the short videos about Beijing, Xi'an, and Shanghai, and serves as a key entry for understanding the application of contemporary strategies in constructing city images.

Contributions

The research contributions are structured in three dimensions based on the key findings. Firstly, this study probes into communication mechanisms and expressive

logic to expand and deepen existing theories on city image construction. Secondly, it develops a multidimensional analytical framework compatible with the short video context, contributing to methodological innovations in research related to city communication. Thirdly, it offers pragmatically valuable insights into city image propagation, content creation, and platform mechanism analysis. The following section will elaborate on these contributions in dimensions of theory, methodology, and practice.

Theoretical contributions

From a theoretical perspective, this study performs case studies of Beijing, Xi'an, and Shanghai to identify the mechanisms underlying city image construction in short videos, offering a fresh viewpoint for research on city communication and media representation.

First, it broadens the scope of city image studies. Unlike previous research often devoted to state-led narratives and macro-level depictions of mainstream media, this study prioritizes user-generated content and visual micro-strategies, underscoring how city images are actively constructed and recreated as supported by the operational mechanism of short videos. As a result, it shifts the research focus from viewing city images as passive representations of the city to the lived-in perception of viewers.

Next, this study deepens the understanding of structural interactions among multiple actors involved in constructing city images. As noted from the comparative analysis, creators are transformed from mere transmitters of information to proactive participants in shaping city imagery. The form and rhythm of content are constructed by the operational mechanism underlying short videos, while the city itself evolves from a static place to a subject continuously selected, interpreted, and redesigned throughout the expressive process. A dynamic model of city communication, as a

supplement to previous research primarily dedicated to content or audiences, is thus proposed by examining the above structural interactions.

Finally, the roles of technology in city image construction are accentuated.

Representational techniques, such as color, light and shadow, and sound, are essential to the observation and perception of cities, since their technical roles in image construction present how the interplay of technology, aesthetics, and communication rationale shapes the construction of city images on digital platforms. This offers a new theoretical gateway for grasping city communication in the context of digital media.

Methodological contributions

The construction of a multidimensional analytical framework fit for studying city image construction in the short video context furnishes research on city communication in the digital media era with novel tools and practical approaches, thus providing methodological contributions.

First, the proposed framework employs a systematic comparative analysis to elaborate on the expressive strategies of short videos about Beijing, Xi'an, and Shanghai under the same operational rationale and technological environment, highlighting both "similarities within differences" and "differences within similarities." This cross-city, cross-content comparative approach helps to more accurately uncover the selection strategies and cultural motivations that support the construction of city images.

Second, it includes a structured content analysis framework that divides the representational techniques into five major dimensions: color, light and shadow, sound, characters, and post-production. This retains the close reading strengths of traditional textual analysis, and captures the intricacy in audiovisual elements and emotional expression paths within visual content. It aligns well with the fast-paced

and aesthetically driven nature of short videos, and broadens the methodological boundaries of conventional city communication research.

Third, it highlights the significance and value of incorporating creators' perspectives into video analysis. The meticulous analysis of expressive choices, rhythmic organization, and emotional design in the videos investigates the content and means of representation and their underlying rationale in depth. This refreshing perspective strengthens the research focus on the relationship between creators' expressive motivations and the short video mechanism, facilitating the research in city image construction.

Practical contributions

Notably, this study offers practical insights for the comprehension of city communication, and the content creation and operational mechanism of short videos.

First, its systematic comparison of visual styles and expressive strategies in the three cities' short videos supports local governments and cultural institutions with specific visual strategy references for city image promotion. Successful cases from different cities have proven that strategies of accurately seizing a city's unique character, refining its visual features, and aligning them with platform-specific communication rationale are central to enhancing a city's appeal and recognizability.

Second, it outlines how representational techniques, including color, light and shadow, and sound, influence the city image construction by creators. This helps creators shape narrative rhythm and emotional tone more effectively when designing city-related content. Moreover, such findings are particularly relevant for creators of cultural tourism accounts and commercial promotional videos.

Third, it provides platform operators with an analytical framework for examining expressive preferences in city-related content and refining content recommendation strategies. In this way, they can further understand creators' expressive tendencies, and better organize and promote city-focused content, thereby enhancing both the diversity of city image construction and the overall efficiency of city image communication.

Limitations of the research

Although this study offers multi-dimensional explorations of city image construction in short video production from theoretical, methodological, and practical perspectives, it bears certain research limitations.

First, it fundamentally relies upon the short video platform Xiaohongshu for research. Thus, platforms with a broader user base, such as TikTok and Kuaishou, can be incorporated into future research to compare influences on city image construction exerted by varied platform mechanisms and user preferences. This can enable a more comprehensive knowledge of the communication rationale within the broader digital platform ecosystem.

Second, its methodology primarily concerns content analysis and visual interpretation, emphasizing the analysis of creators' strategies and the image construction process. Consequently, future research could introduce mixed method research, such as by carrying out interviews, surveys, and communication experiments, or combining audience's perspectives with platform data to acquire deeper insights into the process of image communication.

Finally, approaches to city image construction keep evolving, following the rise of AI-generated content, immersive media, and virtual spaces. In answer to this trend, future

research could explore the interaction between digital visual culture and city image construction as empowered by emerging technologies, thus being conducive to advancing city image studies in a more forward-looking direction.

Above all, this study, despite holding research limitations, offers tentative perspectives regarding constructing city images in short videos, and suggests room for further exploration and refinement in future research.

Concluding remarks

According to the research findings discussed above, city images are no longer static visuals, but are constantly reconstructed through interactions of time, place, and narrative. Such city imagery presented through short videos is now neither defined by a few authoritative voices, nor created by simply superimposing perspectives from conventional promotional films. As shown through countless short videos, creators use the camera lens as a figurative brush to outline new forms of expression.

Technological advancements empower cities to be portrayed with greater clarity and experienced in more immersive ways. The on-screen movements no longer point to the simple flow of information but activate points of connection between the city and viewers. The city image in short videos is more like an invitation written by creators to the world, quietly unfolding with every click.

In all, this study focuses on how cities are represented, specifically how they are narrated, perceived, and imagined. It is the joint efforts of creators and short video platforms that allow each city to transcend geographical boundaries and join global communication networks, during which the city emerges as a collectively shaped and perceived interface. With the ongoing emergence of digital trends, the construction of city images will keep encountering new challenges and possibilities, and short video platforms are rising as key spaces for contemporary creative expression in this

context.

References

- Anholt, S. (2007). *Competitive identity: The new brand management for nations, cities and regions*. Basingstoke: Palgrave Macmillan.
- Anholt, S. (2016). *Places: Identity, image and reputation*. Springer.
- Attademo, G. (2021). Color and/is narration. The narrative role of color in Wes Anderson's filmic images. *Cultura e Scienza del Colore—Color Culture and Science*, 13(01), 07–13.
- Baker, S., Istvandy, L., & Nowak, R. (2016). Curating popular music heritage: storytelling and narrative engagement in popular music museums and exhibitions. *Museum Management and Curatorship*, 31(4), 369-385.
- Barns, S. (2019). *Platform urbanism: Negotiating platform ecosystems in connected cities*. Springer Nature.
- Berland, J. (1993). Sound, image and social space: music video and media reconstruction. *Sound and vision: The music video reader*, 25-43.
- Bezdek, M. A., & Gerrig, R. J. (2008). Musical emotions in the context of narrative film. *Behavioral and Brain Sciences*, 31(5), 578-578.
- Block, B. (2020). *The visual story: Creating the visual structure of film, TV, and digital media*. Routledge.
- Boltz, M., Schulkind, M., & Kantra, S. (1991). Effects of background music on the remembering of filmed events. *Memory & cognition*, 19, 593-606.
- Bordwell, D., Thompson, K., & Smith, J. (2004). *Film art: An introduction* (Vol. 7, pp. 477–479). New York: McGraw-Hill.
- Brennan, K. G., & Pearlman, K. (2023). Creating character in editing. *Media Practice and Education*, 24(3), 235–252.
- Butt, D., McQuire, S., & Papastergiadis, N. (2016). Platforms and public participation. *Continuum*, 30(6), 734-743.
- Cakmak, M. C., Shaik, M., & Agarwal, N. (2024, April). Emotion assessment of youtube videos using color theory. In *Proceedings of the 2024 9th International Conference on Multimedia and Image Processing* (pp. 6–14).

- Cao, L. (2025). Color in the documentary film. *Journal of Humanities, Arts and Social Science*, 9(1).
- Cao, Z. (2021). *Research on communication strategy of short video blogger on tiktok platform*. *International Journal of Social Science and Education Research*, 4(7), 42.
- Caprotti, F., Chang, I. C. C., & Joss, S. (2022). Beyond the smart city: A typology of platform urbanism. *Urban Transformations*, 4(1), 4.
- Chen, H. Y. (2023). A brief discussion on sound creation of short videos in the context of new media (浅谈新媒体背景下短视频的声音创作). *TV Research (电视研究)*, (04), 75-76.
- Chen, I. P., Wu, F. Y., & Lin, C. H. (2012). Characteristic color use in different film genres. *Empirical Studies of the Arts*, 30(1), 39-57.
- Chen, S. B., Wei, J., & Chen, C. Y. (2022). Viewing the city through media: Exploring the impact mechanism of media representation on urban identity under the background of smart city strategy (透过媒介看城市: 智慧城市战略背景下媒介呈现对城市认同的影响机制探究). *Modern Advertising (现代广告)*, (14), 4-16+61.
- Chen, S., & Chu, Y. (2022, December). A study of the cultural significance of traditional colors in Chinese movies. In *2022 6th International Seminar on Education, Management and Social Sciences (ISEMSS 2022)* (pp. 1362-1368). Atlantis Press.
- Chen, Y. (2009). Media construction of urban image: Conceptual analysis and theoretical framework (城市形象的媒体建构——概念分析与理论框架). *Journalism Tribune (新闻界)*, (05), 103-104, 118.
- Cohen, J. (2001). *Visual color and color mixture: the fundamental color space*. University of Illinois Press.
- Cohen, J. (2013). Audience identification with media characters. In *Psychology of entertainment* (pp. 183-197). Routledge.
- Coplan, A. (2008). Empathy and character engagement. In *The Routledge companion to philosophy and film* (pp. 97-110). Routledge.

- Cutting, J. E., Brunick, K. L., DeLong, J. E., Iricinski, C., & Candan, A. (2011). Quicker, faster, darker: Changes in Hollywood film over 75 years. *i-Perception*, 2(6), 569–576.
- Dancyger, K. (2018). *The technique of film and video editing: history, theory, and practice*. Routledge.
- DeNora, T. (2000). *Music in everyday life*. Cambridge University Press.
- Du, X., Liechty, T., Santos, C. A., & Park, J. (2022). ‘I want to record and share my wonderful journey’: Chinese Millennials’ production and sharing of short-form travel videos on TikTok or Douyin. *Current Issues in Tourism*, 25(21), 3412–3424.
- Duan, P. (2022). The visual turn in music communication in the era of short videos (论短视频时代音乐传播的视觉化转向). *Modern Publishing* (现代出版), 35(1), 45-49.
- Fan, X. (2023). The roles of narration, music, and sound effects in documentaries (解说及音乐音效在纪录片中的作用). *Film & Television Production* (影视制作), (09), 89-92.
- Felleman, S. (2014). *Real objects in unreal situations: Modern art in fiction films*. Intellect.
- Feng, D. W. (2019). Analyzing multimodal Chinese discourse: Integrating social semiotic and conceptual metaphor theories. In *The Routledge handbook of Chinese discourse analysis* (pp. 65–81). Routledge.
- Fenwick, S. V. (2024). Symbolic use of color in narrative montage and its impact on temporal and spatial transitions. *Art and Society*, 3(4), 13–19.
- Figueiredo, C. (2015). Narrative and visual cinematic strategies in communication design. *Procedia Manufacturing*, 3(1), 1725–1732.
- Flueckiger, B. (2016). Color and subjectivity in film. In *Subjectivity across Media* (pp. 145-161). Routledge.
- Ford, L. (1994). Sunshine and shadow: lighting and color in the depiction of cities on film. *Place, power, situation and spectacle: A geography of film*, 119–136.

- Ford, L. (2020). "Sunshine and shadow: Lighting and color in the depiction of cities on film" from S. Aitken and L. Zonn (eds), *Place, Power, Situation, and Spectacle* (1994). In *The Urban Geography Reader* (pp. 325–333). Routledge.
- Geng, S., Zhang, J., Fu, Z., Gao, P., Zhang, H., & de Melo, G. (2020). Character matters: Video story understanding with character-aware relations. *arXiv preprint arXiv:2005.08646*.
- Goldie, P. (2003). Narrative, emotion, and perspective. In *Imagination, Philosophy and the arts* (pp. 55–69). Routledge.
- Görne, T. (2019). The emotional impact of sound: A short theory of film sound design. *EPiC Series in Technology*, 1, 17-30.
- Gou, L. (2025). The narrative construction of auditory elements in Chinese films in recent years (近年来我国电影中听觉元素的叙事建构). *China Film News* (中国电影报), p. 011.
- Grodal, T. (2005). Film lighting and mood. *Moving image theory: Ecological considerations*, 152.
- Guo, J. (2024, November). Research on digital narratives of Chinese cultural communication from a cross-cultural perspective. In *Proceedings of the 2024 3rd International Conference on Comprehensive Art and Cultural Communication (CACC 2024)* (Vol. 863, p. 46). Springer Nature.
- Hao, X., & Ryan, C. (2013). Interpretation, film language and tourist destinations: A case study of Hibiscus Town, China. *Annals of tourism research*, 42, 334–358.
- Harvey, D. (2002). *Spaces of capital: Towards a critical geography*. Routledge.
- He Guoping (2010). Urban image communication: Framework and strategy (城市形象传播:框架与策略). *Modern Communication* (现代传播), (08), 13–17.
- Holcomb, B. (2021). Place marketing: Using media to promote cities. In *Imaging the City* (pp. 33–55). Routledge.
- Hopkins, R. (2008). What do we see in film?. *The Journal of Aesthetics and Art Criticism*, 66(2), 149–159.
- Horesh, N. (2014). *Shanghai, past and present: A concise socio-economic history*,

- 1842–2012. Liverpool University Press.
- Huang, R. (2021). The creative expression of light in film narrative (光线在电影叙事中的创造性表达). *Theatre Family* (戏剧之家), (34), 143–145.
- Jing, Z. (2024). A study on the relationship between characters and plot in converged media film and television creation (融媒体影视创作中人物和情节关系研究). *China Newspaper Industry* (中国报业), (12), 66–67.
- Katz, E. (2001). Media events: The sense of occasion. *Studies in Visual Communication*, 10(2), 84–89.
- Kaul, V. (2012). The changing world of media & communication. *Journal of Mass Communication & Journalism*, 2, 116.
- Kaul, V. (2014). Representation of social issues in films. *Madhya Pradesh Journal of Social Sciences*, 19(1), 139–158.
- Kavaratzis, M., & Hatch, M. J. (2013). The dynamics of place brands: An identity-based approach to place branding theory. *Marketing theory*, 13(1), 69–86.
- Ke, H. (2024). Cross-cultural communication strategies for Chinese urban image (中国城市形象的跨文化传播策略). *China Development Review* (中国发展观察), (06), 102–107.
- Keating, P. (2022). Light and time in the narrative fiction film. *JCMS: Journal of Cinema and Media Studies*, 61(3), 59–81.
- Kim, Y. S., & Jun, Y. S. (2010). The images of Chinese traditional colors and cultural preferences-focus on the movie costumes of< The Last Emperor>-. *Journal of the Korean Society of Clothing and Textiles*, 34(12), 2006–2021.
- LaBelle, B. (2019). *Acoustic territories: Sound culture and everyday life*. Bloomsbury Publishing USA.
- Lacey, N. (1998). *Image and representation key concepts in media studies* (1st ed. 1998.). Macmillan Education UK.
- Lavrenova, O. (2023, May). Color semantics of the cultural landscape. In *Arts* (Vol. 12, No. 3, p. 111). MDPI.
- LI, M. (2023). Dialogue or monologue: a study on public participation in urban image communication based on social media. In *Proceedings of the World*

- Conference on Media and Mass Communication* (Vol. 7, No. 01, pp. 39–52).
- Li, M. (2023). Research on Chinese film industry under the background of globalization—Taking Zhang Yimou's film as an example. *Frontiers in Art Research*, 5(14).
- Li, Z. (2025). The "Other" perspective of social networks and cross-cultural imagination: A grounded research based on the audience of "foreign internet celebrities" (社交网络的“他者”视景与跨文化想象-基于“洋网红”受众的扎根研究). *Southeast Communication* (东南传播), 01, 78–81.
- Liao, Y. (2014). *Food and festivals of China*. Simon and Schuster.
- Liu, J., Zhao, J., Wei, W., & Yang, T. (2025). Captivating the crowd: unraveling sentiments in tourism short videos for effective destination marketing on social media. *Current Issues in Tourism*, 1–28.
- Liu, X. M., Luo, Y., Zeng, C. R., Wei, Y. K., Chen, Y. L., & Li, K. X. (2024). The impact of online literary works on college students' travel intentions: An analysis based on plot narration, character images, and work types (网络文学作品对大学生出游意愿的影响-基于情节叙述、人物形象、作品类型的分析). *Journal of Xi'an Shiyou University* (西安石油大学学报), (03), 58–65.
- Liu, X., Wu, J., & Zhao, B. (2022). Research on the impact of content fragmentation propagation in short videos. In *2021 International Conference on Social Development and Media Communication (SDMC 2021)* (pp. 231–236). Atlantis Press.
- Lynch, K. (1964). *The image of the city*. MIT press.
- Magliano, J. P., & Zacks, J. M. (2011). The impact of continuity editing in narrative film on event segmentation. *Cognitive science*, 35(8), 1489–1517.
- Malkiewicz, K. (2012). *Film lighting: Talks with Hollywood's cinematographers and gaffer*. Simon and Schuster.
- Matamala, A. (2018). Voice-over: practice, research and future prospects. In *The Routledge handbook of audiovisual translation* (pp. 64-81). Routledge.
- McClellan, S. T. (2007). *Digital storytelling: The narrative power of visual effects in film*. MIT Press.

- Mei, M. L., & Zhu, Z. Q. (2023). The integration of local and international urban spirit: The construction of Shanghai's urban image through architectural landscape from a spatial perspective (本土化与国际化的城市精神融合: 空间视域下建筑景观对上海城市形象的建构). *Audio-Visual (视听)*, (12), 146–150.
- Meng, X. (2024). Reflection and reconstruction of reality in film art. *Journal of Humanities, Arts and Social Science*, 8(7).
- Ng, M. K., Zhai, B., Zhao, Y., & Li, J. (2016). Valuing xi'an: A Chinese capital city for 13 dynasties. *DisP-The Planning Review*, 52(3), 8–15.
- Oatley, K. (1995). A taxonomy of the emotions of literary response and a theory of identification in fictional narrative. *Poetics*, 23(1-2), 53–74.
- Pavlović, I., & Marković, S. (2011). The effect of music background on the emotional appraisal of film sequences. *Psihologija*, 44(1), 71-91.
- Pearlman, K. (2012). *Cutting rhythms: Shaping the film edit*. Routledge.
- Peng, H. X., & Du, Z. X. (2024). AI creativity and urban communication: A study of media practice based on Shenzhen's "Read Special" (AI 创意与城市传播: 基于对深圳“读特”的媒介实践研究). *Era of Convergent Media (融媒时代)*, (12), 4-9.
- Peng, Jixiang (2018). *Film and television aesthetics(影视美学)*. Peking University Press(北京大学出版社).
- Plantinga, C. (2008). Emotion and affect. In *The Routledge companion to philosophy and film* (pp. 86–96). Routledge.
- Qian, X. (2022). Three narrative patterns of the city image visually presented on Instagram under the influence of self-presentation. *Media, Culture & Society*, 44(6), 1149–1165.
- Relph, E. (1976). *Place and placelessness* (Vol. 67, p. 45). London: Pion.
- Rodgers, S. (2014). The architectures of media power: Editing, the newsroom, and urban public space. *Space and Culture*, 17(1), 21–35.

- Säks, E. (2023). Lighting in cinematic virtual reality. *International Journal on Stereo & Immersive Media*, 7(1), 122–135.
- Shao, W. Y., & Ni, X. B. (2018). A new approach in documentaries: First-person voice-over narration (纪录片新方法: 第一人称画外音叙事). *Modern Communication (现代传播)*, (03), 110-113.
- Shiping, G. (2025). The significant form: Visual symbols and cultural symbolism in film. *Frontiers in Art Research*, 7(1).
- Smith, T., & Henderson, J. (2008). Attentional synchrony in static and dynamic scenes. *Journal of Vision*, 8(6), 773–773.
- Spirn, A. W. (1988). The poetics of city and nature: Towards a new aesthetic for urban design. *Landscape Journal*, 7(2), 108–126.
- Tseng, C. I. (2013). Analysing characters' interactions in filmic text: a functional semiotic approach. *Social Semiotics*, 23(5), 587–605.
- Van Dijck, J., Poell, T., & De Waal, M. (2018). *The platform society: Public values in a connective world*. Oxford university press.
- Van Leeuwen, T., & Boeriis, M. (2016). Towards a semiotics of film lighting. In *Film Text Analysis* (pp. 24–45). Routledge.
- Wang, H. T. (2024). A narrative study of spatial montage in the era of digital special effects (数字特效时代空间蒙太奇的叙事研究). *Movie Literature (电影文学)*, (12), 129–134.
- Wang, H., Yang, S., & Dai, Z. (2022). The development of mobile short video communication in the context of the mobile internet. *Science Insights*, 40(1), 421–426.
- Wang, J. (2011). *Beijing record: A physical and political history of planning modern Beijing*. World Scientific.
- Wang, J. R. (2022). The expansion of film visual space through light digitization (光线数字化对电影视觉空间的拓展). *Contemporary Cinema (当代电影)*, (4), 171–176.
- Wang, M. (2025). The reshaping and innovation of film narrative, rhythm, and editing in the era of fragmented viewing(碎片化观影时代电影叙事、节奏、剪辑的

- 重塑与创新), *Movie Literature* (电影文学), (01), 103–107.
- Wang, Q. (2024). Light and shadow: exploring the aesthetics of *Dune*'s visual spectacle through effects, design, and art direction. *Journal of Research in Social Science and Humanities*, 3(2), 14–21.
- Wang, X. (2023). Urban communication strategy based on short video platform: A case study of TikTok. In *SHS Web of Conferences* (Vol. 155, p. 02018). EDP Sciences.
- Wang, Y. (2013). The charm of form: Space and imagery in film colors (形式的魅力: 电影色彩里的空间与意象). *Movie Literature* (电影文学), (23), 147–148.
- Wang, Y. N. (2013). The influence of color and light in micro-films on audience emotions (微电影中的色彩与光线对观众情绪的影响). *Journalism World* (新闻世界), (7), 277–279.
- Watkins, Liz. (2002). Light, colour and sound in cinema. *Paragraph*, 25(3), 117–128.
- Wei, W. (2020, May). The construction of Chengdu city image in new media environment: take TikTok as an example. In *6th International Conference on Humanities and Social Science Research (ICHSSR 2020)* (pp. 477–480). Atlantis Press.
- Wisessing, P., Zibrek, K., Cunningham, D. W., Dingliana, J., & McDonnell, R. (2020). Enlighten me: Importance of brightness and shadow for character emotion and appeal. *ACM Transactions on Graphics (TOG)*, 39(3), 1–12.
- Wu, M., & Feng, Y. (2021). The context of color: A study on the visual color of the city—Taking Chengdu as an example. *Art and Design Review*, 10(1), 29–40.
- Xi, L. Y. (2024). Urban spatial micro-narratives and path reflection: The shaping of urban image in short videos (城市空间微叙事与路径反思: 短视频中的城市形象塑造). *Editor's Friend* (编辑之友), 2024(11), 66–72.
- Xing, W., Lv, B., Liu, T., & Jiang, L. (2024). The evolution of China's opening-up policy and the measurement of the opening-up level of major coastal cities. *Marine Development*, 2(1), 16.
- Xu, Qingmiao. "Shunji Iwai's cinematic narrative." *Studies in Art and Architecture* 2, no. 4 (2023): 58–62.

- Yang, C., Zhao, Y., Zheng, H., & Zhu, Z. (2024). External communication effect of short video of Chinese We-Media—Take the topic of Chinese food culture on YOUTUBE as an example. *Lecture Notes in Education Psychology and Public Media*, 56, 157–164.
- Yao, Y., & Zhuo, Y. (2018). A multimodal discourse analysis of the promotional video of Hangzhou. *English Language Teaching*, 11(10), 121–131.
- Yeh, E. Y. Y., & Hu, L. W. (2008). Transcultural sounds: Music, identity, and the cinema of Wong Kar-wai. *Asian Cinema*, 19(1), 32-46.
- Yin, J., & Gan, T. (2025). Empowerment of community narrative: Characteristics and strategies of urban image communication in the context of new media (共同体叙述赋能：新媒体语境下城市形象传播的特征与策略). *Journal of Ethnology* (民族学刊), (02), 57–66, 146.
- Yu, J. (2023). A Study on the narrative space of rural short video: A case study of rural short video in TikTok platform. *Frontiers in Art Research*, 5(2).
- Zhan, H. (2024). Aesthetic characteristics and creative practice of micro-films from the perspective of directors. *Frontiers in Art Research*, 6(11).
- Zhang, Ailing. (2024). The symbolic construction of urban images in short videos featuring intangible cultural heritage (非遗类短视频中城市形象的符号建构). *Audiovisual Horizon* (视听界), 5, 107–108.
- Zhang, K. (2024). Empathy and arousal in the construction of Beijing's urban image: Taking the series of works on the Beijing central axis as an example (北京城市形象建构的共情与唤起-以北京中轴线系列作品为例). *Sound and Screen World* (声屏世界), (16), 42–44.
- Zhang, X. N. (2022). The visual art aesthetics in Zhang Yimou's *Shadow* (张艺谋《影》中的视觉艺术美学). *Media Forum* (传媒论坛), (10), 66–68.
- Zhao Yue. (2024). A decade of breaking boundaries: Xiaohongshu's breakthrough and dilemmas in the era of intelligent media (十年破圈之路：智媒时代小红书的突围与困顿). *Sound & Screen World* (声屏世界). (08), 86–88, 110.

- Zhao, Y. (2021). Reflections on the application of light and space in film photography styling (电影摄影造型中光线和空间运用思考). *Daguan* (大观), (10), 60–61.
- Zhou, Xian (2017). Contemporary Chinese visual culture studies (当代中国的视觉文化研究). Yilin Press (译林出版社).
- Zou, S. (2025). Transcoding a wanghong city: Mediatized culturalization of urban places in China. *Media, Culture & Society*, 47(1), 22–40.
- Zukin, S. (2009). *Naked city: The death and life of authentic urban places*. Oxford University Press.

Appendix 1 Video List

Beijing:

| No. | Account | Video title | Publication date |
|-----|-----------------------------------|---|------------------|
| 1 | Fangqi Kiki | The Red Walls and Green Tiles of the Forbidden City Have Locked Away the Youth of Countless Women | 25.1.2020 |
| 2 | Lili Shanhe | You Have to Climb the Great Wall Once: Feeding Cats, Eating Popsicles, and Admiring the Vast Mountains and Rivers | 28.2.2023 |
| 3 | Jianchuan Silent | Beneath the Great Wall, Intangible Cultural Heritage Shines Through Time | 3.5.2023 |
| 4 | Xiwai CYAN | Listen to the Voice of This City in Stillness | 19.9.2020 |
| 5 | Wang Mengyun | Beijing: Bei Stands for Beiping, and Jing Stands for Capital | 29.4.2023 |
| 6 | If the Camera Could Speak | At This Moment, Beijing Is the 'Paradise on Earth' Once Praised by Lao She | 3.11.2022 |
| 7 | Mili Meliving | Street Observer: Beijing Citywalk, Life Is Nowhere Else | 4.8.2023 |
| 8 | Jiuwei Ya | I Declare Beijing's Most Romantic Season Is Coming Soon | 10.2.2023 |
| 9 | Xijiang Yue | Cinematic Feel Beijing's Autumn Is Heaven | 13.10.2023 |
| 10 | Fashion Tourism | Most Beijing | 15.7.2023 |
| 11 | Kuntao | Winter of Beiping | 4.2.2024 |
| 12 | sammi | Snowy Hutongs Hold the Warmest Vitality of Life | 13.12.2023 |
| 13 | Heize Kurosawa | Don't Go to Beijing | 16.10.2023 |
| 14 | Tianya LG | Who Gets It I Really Love This Vibe | 26.1.2024 |
| 15 | Xiaoge Shanyue | Winter Shichahai Don't Miss the Unique New Year Flavor of Beijing | 2.2.2024 |
| 16 | Seeing the World Through the Lens | Beijing's New Year Flavor Filmed on Qianmen Street Like a Spring Festival Movie | 27.1.2024 |
| 17 | Yuguan Tianxia | Will Your First Stop in 2024 Be the Capital Beijing | 3.1.2024 |
| 18 | Tang Gaole | This Is What Spring in Beijing's Hutongs Should Look Like | 3.5.2022 |

| | | | |
|----|-----------------------------------|---|------------|
| 19 | Niguang Ji | After Traveling to 90 Countries I Realized My Home Is Still the Most City | 15.8.2024 |
| 20 | Mili Meliving | Beijing Citywalk Lao She Said This Is the Most Beautiful Street in Beijing | 8.10.2023 |
| 21 | New Power of Chinese Animation | Chaoyang District “14th Five-Year Plan” Warm video | 5.1.2022 |
| 22 | Dong Xinwen | AI My Beijing Twelve Data Points Reveal the Ecological Answer Sheet | 21.4.2024 |
| 23 | Seeing the World Through the Lens | Feel the Festive Atmosphere with Lanterns and Decorations at Qianmen Beijingsfang | 7.2.2024 |
| 24 | Daily Advertising Cases | In Tune with the World Here | 16.12.2022 |
| 25 | Mati Cunzhuang | In This Winter You Have to See the Golden Light Piercing the Seventeen-Arch Bridge in the Summer Palace | 25.11.2024 |
| 26 | zhangdarlin | In Winter Beijing I Hear the Ancient Dragon’s Roar | 16.1.2024 |
| 27 | Tianya LG | Finally Understand Why Everyone Has an Obsession with Beijing | 20.6.2024 |
| 28 | Director Wang Zuo | You Might Not Understand the Trendiness of Haidian | 27.7.2024 |
| 29 | Audi | Beijing Drive Walk This Time the Drive Tops It All | 23.11.2023 |
| 30 | Xiaopeng Motors | Bet You’ve Never Seen Beijing’s Friday Rush Hour | 27.12.2023 |
| 31 | Mili Meliving | Old Beijing New Year Is Absolutely Amazing | 28.8.2025 |
| 32 | Seeing the World Through the Lens | Click In to Feel the Festive Firework Atmosphere of Qianmen | 6.2.2024 |
| 33 | Mili Meliving | Auspicious Stars Shine on the Chinese New Year | 31.1.2025 |

Xi’an:

| No. | Account | Video title | Publication date |
|-----|---------|------------------------------|------------------|
| 1 | XiaMu | Don’t Doubt It This Is Xi’an | 23.11.2023 |

| | | | |
|----|--|--|------------|
| 2 | XiaMu | When All the Snow of Chang'an Has Fallen No Dream Person Remains from the Past | 20.11.2023 |
| 3 | Fangqi Kiki | Longing for Each Other in Chang'an | 24.4.2023 |
| 4 | Wang Napao | A Thousand People Singing "Later" at Xi'an Bell Tower | 16.3.2023 |
| 5 | Lili Shanhe | When You Talk About Xi'an You Can't Just Talk About Xi'an | 25.2.2023 |
| 6 | Fashion Tourism | When Night Falls Xi'an Turns into Chang'an | 7.8.2023 |
| 7 | Mada Fa | Late Summer and Early Autumn Cinematic Streets of Xi'an | 5.9.2023 |
| 8 | Don't Know What to Film | Xi'an, Please Take Another Look – A VLOG About This City | 3.4.2023 |
| 9 | JING Visual Studio | Best Viewing Season The Ginkgo King Planted by Li Shimin in Xi'an Is Stunning | 8.11.2023 |
| 10 | Qian_Ma | I Want to Accompany You from Morning Bell to Evening Drum | 16.6.2023 |
| 11 | Star and Moon Moments | The Only City Worthy of the Name of the Prosperous Tang Is Chang'an | 14.4.2023 |
| 12 | Chahuanong | This Is Chang'an Hello Chang'an | 29.4.2023 |
| 13 | A Blue Whale | Tonight the Light Shines for Thousands of Miles No City Is More Eye-Catching than Chang'an | 2.1.2023 |
| 14 | liandezhi | Night Tour in Chang'an City A Toast to the Tang Dynasty and to Delicious Food | 24.3.2023 |
| 15 | Shi Yihan | Chang'an, Ever Safe | 22.6.2023 |
| 16 | Wind from the Southern Foothills | Snowy Day, Warring States Robes, and a Full Sense of Destiny | 17.1.2024 |
| 17 | Dong Dongzi | You Must Visit Xi'an the City of Carbs | 18.8.2024 |
| 18 | E S | Tonight the Xi'an City Wall Spring Festival Lantern Fair Officially Lights Up | 2.2.2024 |
| 19 | Xi'an Travel – Xiaofu | Shocked for the First Time by the In-Flight Announcement in Xi'an | 6.7.2024 |
| 20 | Romantic Little Sheep | This Is Xi'an Also Chang'an | 24.2.2024 |
| 21 | Cass LV | Once You Enter Chang'an You Gaze upon a Thousand Years | 17.10.2024 |
| 22 | Xiao'an's Street Diary | I Declare This Section of Xi'an's City Wall Is Absolutely Legendary | 8.9.2024 |
| 23 | Hu Yi | A Video of Hanfu Restoration | 6.1.2020 |

| | | | |
|----|----------------------------------|---|------------|
| 24 | Mr. Crab | Xi'an in My Eyes | 20.9.2020 |
| 25 | Super Hengyun | Under the City Wall of Xi'an Runs the Train of Xi'an People | 19.9.2023 |
| 26 | Rou Rou Taichai | Life Advice Experience Xi'an's Morning Market at Least Once | 22.1.2024 |
| 27 | Ultimate Aerial Photographer DFG | The Central Axis of Xi'an Turns Out to Be So Beautiful | 2.4.2024 |
| 28 | Super Lynn | If You Have Only One Day in Xi'an Visit This Route before You Leave | 14.3.2024 |
| 29 | It's Yezi | A Tribute to This Encounter | 11.7.2022 |
| 30 | Nic Paipaikan Ba | This Is How You Should Edit Your Travel Video in Xi'an | 29.11.2023 |
| 31 | Mu Qianjie | Xi'an City Wall | 29.7.2024 |
| 32 | Sun Daye | One of Only Five Sofitel Legend Hotels in the World and the Only One in China | 17.4.2023 |
| 33 | A Gentle Pat on Xi'an | A Thousand-Year-Old Meipi and the Snowy Qinling Mountains in One Stunning Frame | 28.11.2024 |
| 34 | Photographer Zebra | Come to Xi'an to Film in Warring States Robes | 4.5.2025 |
| 35 | Chef Guangtan | French Chef Travels in Xi'an | 28.11.2021 |

Shanghai:

| No. | Account | Video title | Publication date |
|-----|---------------|---|------------------|
| 1 | Fangqi Kiki | Riding the Sunset Rollercoaster in Summer Watching a Tango in the Magic City | 20.7.2023 |
| 2 | Fangqi Kiki | Have You Ever Taken a Double-Decker Bus in Shanghai This Is a City That Never Sleeps | 4.8.2021 |
| 3 | Lili Shanhe | Absolutely Love the Charm of Shanghai Where Prosperity Meets Civilization | 23.4.2023 |
| 4 | Wushiyou Rain | In Shanghai Searching for the Memories of Silent Separation from Seven Years Ago | 19.11.2022 |
| 5 | Wushiyou Rain | I've Decided to Go to Shanghai to Find the Filming Locations of Dear, Loved | 2.12.2022 |
| 6 | Wushiyou Rain | Unable to Escape Ming and Xun's Fate I Came to Shanghai to Find the Filming Locations of Light Up My Warmth for You | 13.12.2022 |
| 7 | Wushiyou Rain | After Several Rounds around the Bund I Summarized a Night View Guide for the Sightseeing Bus | 16.9.2022 |

| | | | |
|----|-------------------------------|---|------------|
| 8 | Wushiyu Rain | After Ten Years I Still Miss the Shanghai Depicted in Tiny Times | 14.9.2022 |
| 9 | Shaoyuelun Allen | Has Autumn Arrived in Your City | 19.10.2021 |
| 10 | Shaoyuelun Allen | Plane Trees Old Houses and Narrow Alleys Hide a Piece of the Past | 5.7.2021 |
| 11 | Xi Yuexue | Half of Shanghai's Romance Lies on Wukang Road | 9.5.2023 |
| 12 | Almond | You Can Always Believe in Shanghai's Autumn | 28.11.2021 |
| 13 | Dreamer Jiang Shiyi | What Kind of City Do You Think Shanghai Is | 20.3.2021 |
| 14 | Fashion Tourism | Strolling through the Shanghai in Wong Kar-wai's Eyes | 11.1.2024 |
| 15 | Fashion Tourism | Old Times in Shanghai's Alleys | 10.7.2023 |
| 16 | Haodong Tongxue | Aerial View of the Glamorous and Prosperous Night Scene of the Bund | 1.3.2024 |
| 17 | zhangdarlin | I Use a Cinematic Approach to Capture the Old Times of Shanghai | 24.1.2024 |
| 18 | Xiaoge Shanyue | Spring Train to Gongqing Park Wandering into Monet's Garden | 15.3.2024 |
| 19 | REDclub | A Giant Potato Appeared in Front of Xiaohongshu's Entrance | 16.10.2024 |
| 20 | A Moment in the Magic City | A Famous Shanghai Landmark More Stunning than the Bund Yet Overlooked by Tourists | 10.10.2023 |
| 21 | Romantic Little Sheep | Exploring Shanghai in 24 Hours from Glamour to Plane Tree Avenues | 14.5.2023 |
| 22 | Where Did Muffin Go Today | Promise Me You'll Spend Christmas in Shanghai This Year | 4.11.2024 |
| 23 | Liuxiao Bamboo | This Half Minute Captures the Taste of Shanghai | 19.2.2024 |
| 24 | Huyi | Where Are You This Winter How Have You Been | 2.12.2020 |
| 25 | zhangdarlin | A Wonderful Tipsy Moment Surrender Yourself to the Night | 5.9.2022 |
| 26 | Ashin's Little Park | Time Stumbles Forward as Seasons Come and Go | 19.3.2024 |
| 27 | Ling with Three Dots of Water | How to Explore Inside the Shanghai Bund's International Architecture Complex | 17.12.2024 |

| | | | |
|----|-------------------------------|--|-----------|
| 28 | Duola Dimeng | In April on Earth May We Return to the Beauty of the Past | 9.3.2022 |
| 29 | 17 (Travel After Resignation) | Shanghai One-Day Tour Route Summarized after Walking Thirty Thousand Steps | 4.3.2024 |
| 30 | Louis Vuitton | Enjoying Chinese Culture with Chloë Grace Moretz | 23.4.2024 |