Constructing City Images on Short-video Platforms: A Content Analysis of Douyin

Xu Zhang*
School of Communication, Soochow University, Suzhou 215123, China
*Corresponding email: jacey.zhang@foxmail.com

Abstract

The intelligent algorithm generated content recommendation and the multimodal dissemination of information have brought about changes in the production and dissemination of city images, and short-video platforms have become important media for image dissemination and cultural cognition. To understand the nature of city image construction on the Douyin platform and how to link city image construction to users' engagement, this study took six cities as examples: Beijing, Shanghai, Chengdu, Hangzhou, Xi'an and Chongqing. A multidimensional coding system was designed encompassing content producers, content themes, narrative features, visual symbols and emotional tendencies. Text semantic analysis, visual feature extraction, descriptive statistics, and negative binomial regression were used for the empirical analysis. Based on the analysis of sample data, the following conclusions can be drawn: Lifestyle experience and historical culture are themes of urban image presentation. Ordinary users play a high role in urban image content production. There are differences in the methods of urban image presentation among different content producers; and lifestyle experience themes, experience narration, and positive emotional expression are all positively correlated with user interaction. Accordingly, this study constructs an explanatory framework. It unpacks the process of urban image construction from four dimensions: content selection, symbolic representation, emotional narrative, and interactive feedback. This framework can serve as a methodological reference for quantitative research on urban image dissemination within short-video platforms.

Keywords

Short-video platforms, Douyin, Urban image construction, Content analysis, User interaction

Introduction

Intelligent algorithms have transformed pathways of dissemination of the city image through content recommendation mechanisms. The Douyin platform has reshaped urban space, culture symbols and daily life through multimodal audiovisual expressions and user interaction mechanisms, thus making them important materials for building digital urban images.

In terms of short videos and city image publicity, Li studied the influence of short-video platform on the city image publicity of "new first-tier" cities in China [1]. Wang et al analyzed the mechanism of urban image construction and dissemination in the mobile short-video platform and further studied the influence of user-generated content on urban image dissemination of Tianjin [2,3]. Zhuang and Wang studied the construction and dissemination strategies of urban tourism image in the mobile short-video environment [4]. On the basis of comparative urban research, Liu et al interpreted the construction process of urban spatial image from spatial representation, cultural symbol and online relationship in short-video [5]. Song analyzed the dual effect of regional image construction in Chinese short videos [6]. Zhang et al studied the influence of short-video on destination marketing through comparative research of urban images [7]. Zhao and Zhang studied the visual propaganda attributes of news media in Douyin short video [8]. Cachulo Pereira and Mendes Bacalhau constructed a comprehensive analytical framework for the influence of short videos on destination marketing and local brand communication [9].

Existing studies have examined the relationships among platform communication, spatial representation and urban image formation. However, quantitative analyses covering platform communication, narrative features,

visual symbols and emotional expressions remain incomplete. Moreover, few studies systematically analyze the links between the characteristics of communication content and user interaction. According to this, this study chooses short-video on Douyin from 6 cities, such as Beijing and Shanghai, to be research subjects. A multidimensional content coding system is designed and the contents distribution, differences among actors and interactive relationship are explored based on multimodal feature extraction, descriptive statistics and regression analysis. The purpose of this is to explain the intrinsic logic of urban image construction on short-video platforms and offer an empirical analytical foundation for the research on urban digital communication.

Theoretical foundation and analytical framework for city image construction

Characteristics of city image communication on short-video platforms

The city image dissemination model of short-video platforms is built based on the characteristics of audiovisual integration, algorithmic dissemination, and interactive dissemination. In terms of content presentation, dynamic visuals, background music, subtitles and editing together constitute a symbolic expression system of urban space. Through this system, natural landscapes, historical culture and social life obtain concrete media representation [10].

In terms of dissemination structure, government departments, mainstream media, commercial organizations and ordinary users all participate in the production of communication content. This shifts the construction of urban images from one-way propaganda to pluralistic dissemination. When it comes to dissemination pathways, recommendation algorithms on platforms match information based on user behaviors and information attributes. These algorithms determine the scope and extent of the dissemination of urban image information. Likes, comments and shares can involve the audience in interactive behaviors, which make the city's image present dynamic construction features in the dissemination and negotiation of meanings [11].

Core dimensions of city image construction

The image of a city is the construction of this image, which is built around the objective space of the city, cultural resources and social life of the city as a whole, shaped by the selection of certain media. According to the attributes of short-video content, the image of the city is summarized into five dimensions: natural landscape, city heritage, and culture, modern development, life experience and social and culture [12]. The natural landscape dimension refers to environment, ecology, and spatial character; the historical and cultural dimension is represented by culture, cultural heritage, and memory. The modern development dimension relates to infrastructure and urban development. The lifestyle experience dimension is represented by consumption scenarios and everyday activities, and the social and cultural dimension by public services, citizens' spirit and social behavior.

Each dimension, through visual symbols, narrative content, and emotional expression, collectively constitutes the cognitive structure of the city's image. Its comprehensive representation can be expressed as follows:

$$I_c = \sum_{k=1}^{5} w_k S_k, \quad \sum_{k=1}^{5} w_k = 1 \tag{1}$$

Where I_c is the comprehensive representation index of the city image; S_k is the standardized content representation score for the k th dimension; and w_k is the weight of the corresponding dimension, which can be determined based on coding frequency or preset evaluation rules [13].

Content production and dissemination mechanisms for city image on the Douyin platform

The city image construction on Douyin platform is achieved by content production from different actors, and content is disseminated through symbolic encoding, algorithmic distribution, and user interaction to create a dynamic dissemination mechanism. In the content production, government agencies highlight their achievements in urban development and shaping of public image. Mainstream media emphasize urban events and urban culture values; commercial entities emphasize tourism and consumption scenarios. Ordinary users emphasize their own experiences in urban space through the recording of their daily life [14].

At the level of content expression, urban landmarks, local historic buildings, local cuisine and folk activities are reconstructed into visual symbols. Combined with music, subtitles and emotional narratives, they form a media image of the city with distinctive local identity. On the dissemination level, the platform's recommendation mechanism matches between contents' characteristics

and users' behaviors, which affects the visibility and reach of various types of city image contents. Through behaviors such as liking, commenting and sharing, users participate in content dissemination and meaning negotiation. This intertwines content production, symbolic representation, platform dissemination and interactive feedback. It facilitates the dynamic construction of urban image via a multi-agent expressive system and continuous interaction [15].

Construction of a content analysis framework for urban image

Drawing on the basic dimensions of urban image construction and the communication logic of the Douyin platform, this study builds a multidimensional content analysis framework. It explores six dimensions: communication actors, content themes, narrative features, visual symbols, emotional tendencies, and user interaction. This framework is grounded in urban image theory, media representation theory and local identity theory. It classifies urban image themes into five categories: natural landscape image, historical and cultural heritage image, modern development image, lifestyle image, and social and cultural image. It further identifies the features of urban image content through textual semantics, visual scenes and emotional expression [16].

Building on this foundation, this research takes communication agents, content dissemination features and user interaction metrics as variables. It examines the interrelationships among different content elements. The analysis further identifies the symbolic representations and interactive dissemination characteristics of urban image. The overall analytical framework is shown in Figure 1.

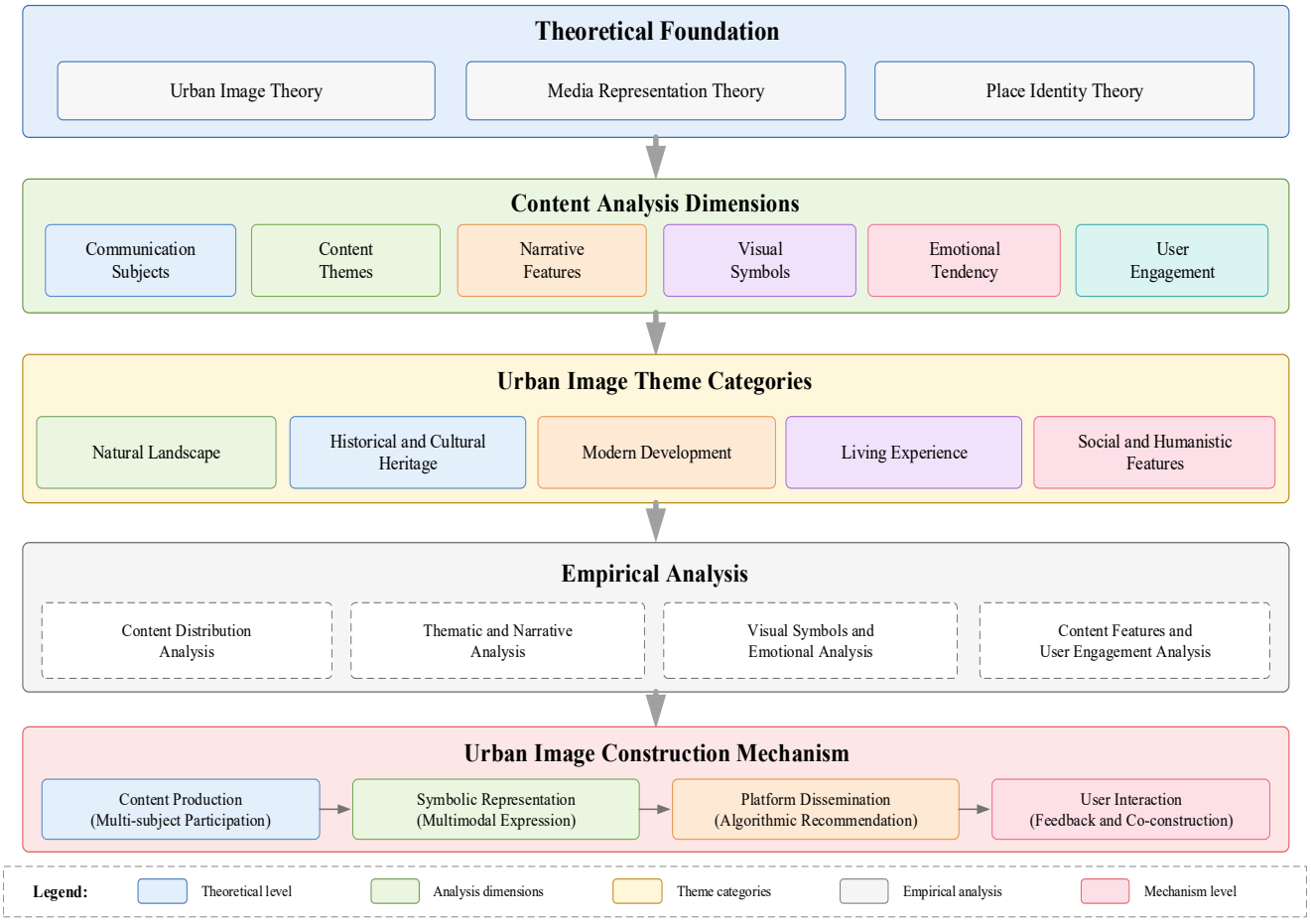


Figure 1. Overall framework diagram.

Research design and data processing

Research subjects and sample selection

The object of this study is short videos of urban image under Douyin platform. Six typical cities were chosen according to their unique development characteristics and cultural location: Beijing, Shanghai, Chengdu, Hangzhou, Xi'an and Chongqing, including two modern cities, two historical and cultural cities, and two tourism and leisure cities.

The study period was determined as January 2025 to

December 2025. A mixed sampling technique - stratified and systematic sampling - was used with search words like city names, urban culture, city tourism, and urban life. Collection of 3, 000 short videos were planned (500 videos/city).

The sample included 4 categories of content creators: government agencies, news media, commercial and general users. Screening criteria were developed to exclude duplicate videos, off-topic videos and videos without major fields, which resulted in a research sample for content coding and multimodal feature analysis.

Collection of Douyin short-video data on city images

Six cities' short-video samples were acquired by public search channels on the Douyin platform. The video title, publication date, duration, publisher, likes, number of comments and number of shares were gathered, as well as subtitle text and keyframe images. A structured database was created with a standardized format of each field and duplicate records, irrelevant videos, and anomalous fields were cleaned [17].

To quantify user interaction levels, a comprehensive interaction index was constructed:

$$E_i = \alpha\, ln(1 + L_i) + \beta\, ln(1 + C_i) + \gamma\, ln(1 + S_i) \quad (2)$$

Where E_i is the comprehensive interaction index for the i th video; L_i, C_i, S_i represent the number of likes, comments, and shares, respectively; α, β, γ are the corresponding weights, satisfying $\alpha + \beta + \gamma = 1$, and were all set to 1 out of 3 in the baseline analysis.

A logarithmic transformation was performed to mitigate the influence of extreme interaction values. A unified data collection time point was also set to lay a solid data foundation for the subsequent analysis of the relationship between content characteristics and user interaction [18].

Construction of a content coding system for city images

A hierarchical coding system was developed to reflect the multidimensional nature of urban image, involving communication agents, content themes, narrative features, visual symbols, emotional tendencies and user interaction [19]. There are four types of communication agents, namely governmental agencies, news media, commercial entities, and users (in Table 1). There are five themes in the contents, namely natural landscapes, history and culture, modern development, lifestyle experiences, and social and cultural aspects. Video expression styles are classified as the characteristics of the narration. Multi-tag coding is used in visual symbols, and positive, neutral and negative are the categories of emotional tendencies.

Table 1. Coding indicator system for urban image content.

Primary indicator	Secondary indicator	Coding rule
Communication sources	Government, Media, Commercial organizations, Individual users	1-4, Single choice
Content themes	Nature, Culture, Development, Daily Life, Humanities	1-5, Single choice
Narrative features	Commentary, Experience, Storytelling, Presentation	1-4, Single choice
Visual symbols	Landmarks, Buildings, Streets, Food, People, Landscapes	0 or 1, Multiple labels
Emotional sentiments	Positive, Neutral, Negative	1-3, Single choice
User engagements	Number of likes, Comments, Shares	Non-negative integers

To quantify the multidimensional characteristics of urban image content, a video encoding vector is constructed:

$$X_i = [T_i, N_i, V_{ij}, \ldots, V_{im}, A_i] \quad (3)$$

Where, X_i represents the content coding vector for the video at index i; T_i, N_i, A_i denote the content theme, narrative type, and sentiment category, respectively; V_{ij} indicates whether the visual symbol in category j appears, with a value of 0 or 1; and m is the number of visual symbol categories [20]. The dissemination agents and interaction metrics are stored as independent analytical variables for subsequent statistical modeling and correlation testing. Sample annotation, data verification, and structured processing were carried out in accordance with the coding manual formulated in advance.

Every coded record was double-checked to reduce human coding bias, and the specific workflow of the whole content analysis procedure is shown in Figure 2.

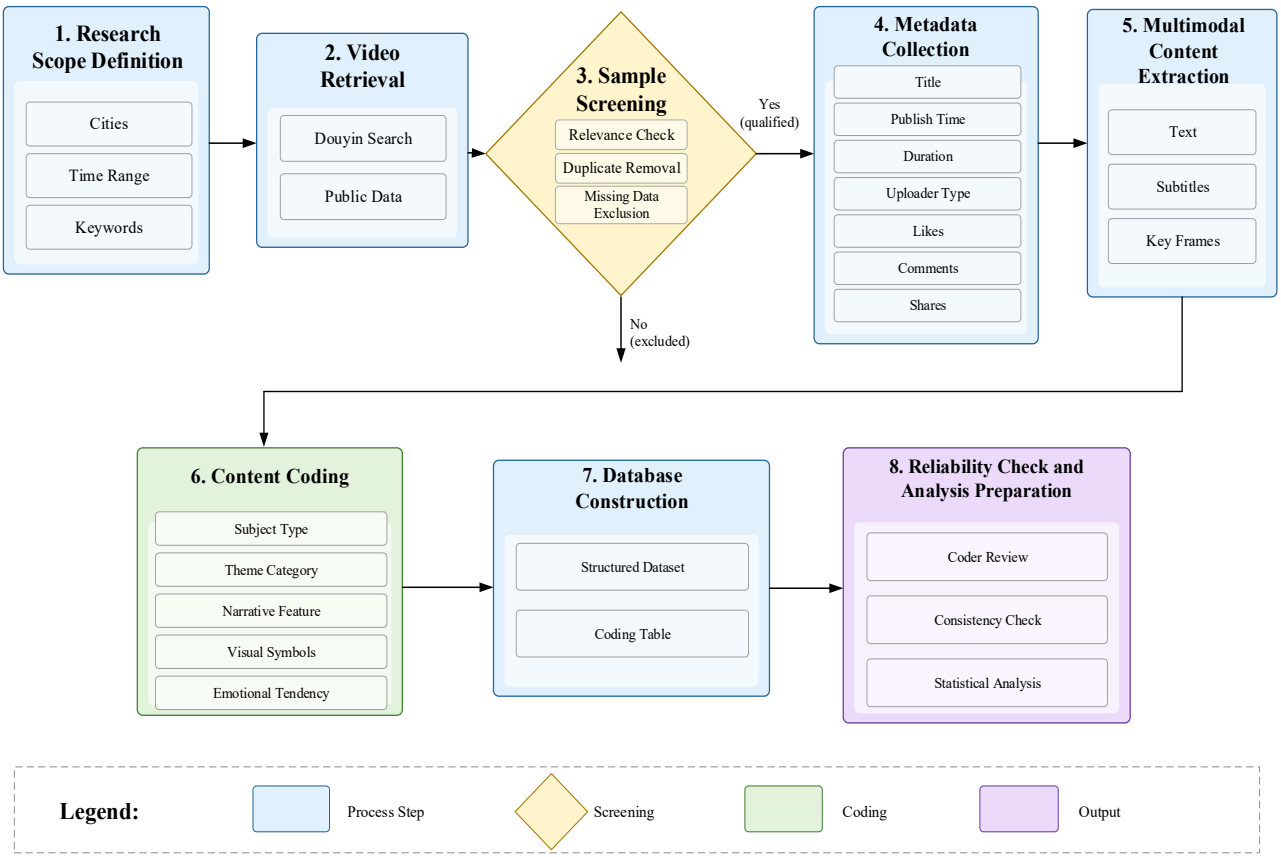


Figure 2. Flowchart of Douyin short-video data collection and content coding.

Multimodal content feature extraction and quantitative analysis

To address the heterogeneous characteristics of text, visual, and emotional information in Douyin short videos, a method for multimodal content feature extraction and quantitative analysis was developed. For the text modality, Chinese word segmentation and the term frequency-inverse document frequency (TF-IDF) algorithm were used to extract keyword features from titles, subtitles, and speech transcripts, combined with semantic classification to identify city image themes and narrative types.

For the visual modality, key frames are extracted at 2-second intervals. Image feature recognition and manual review are used to annotate symbols such as city landmarks, historic buildings, natural landscapes, and everyday scenes. For the sentiment modality, text sentiment scores and audiovisual expression features are integrated to classify content into positive, neutral, and negative sentiments. To achieve multimodal information fusion, a standardized feature representation is constructed:

$$F_i = \lambda_t T_i + \lambda_v V_i + \lambda_e E_i \quad (4)$$

where F_i is the fused feature vector for the video at index i ; T_i, V_i, E_i are the text, visual, and emotional feature vectors, respectively, mapped to a unified feature space [21]. λ_t, λ_v, λ_e are the modal weights, such that the sum of the three equals 1.

Through frequency statistics, cross-analysis, and correlation tests, it quantifies the distribution of urban image content, symbol combinations, and emotional expression characteristics, providing variable support for analyzing differences among communication agents and user interaction relationships.

Reliability and validity testing of content coding

In order to solve this problem, a method of multi-modal feature extraction and quantitative analysis of content in short videos of Douyin was proposed. In terms of text modality, Chinese word segmentation and TF-IDF algorithm were adopted to extract the key words features from title, subtitle and speech transcript text. Combined with semantic classification, this approach identifies the city image themes and the type of the speech, such as positive, negative and neutral.

In terms of the visual modality, the key frames were extracted at 2-second interval, and the image features were recognized by the method of image recognition and

manual check, such as city features, historic buildings, natural scenes, everyday scenes etc. In terms of the sentiment modality, the text sentiment scores and the audiovisual expression features were combined together and classified into positive, neutral and negative sentiments.

Cohen's Kappa coefficient was used to evaluate coding consistency for categorical variables, calculated as follows:

$$\kappa = \frac{P_o - P_e}{1 - P_e} \quad (5)$$

where P_o is the actual agreement rate between the two coders, and P_e is the expected agreement rate under random conditions [22]. A threshold of $\kappa \geqslant 0.80$ was set as the reference standard for high agreement; for indicators falling below this standard, the coding definitions were re-examined, and the coding was re-checked. Validity testing integrates theoretical consistency, alignment with existing literature, expert assessment, and predictive analysis. It evaluates how comprehensively the indicator system covers urban image themes, symbolic expressions, and emotional characteristics. Consistency coefficients for each category were calculated to analyze multi-label visual symbols. Disputed samples were rechecked and coding rules were revised, so as to improve the reliability of content measurement and the validity of analytical results.

Empirical analysis of urban image construction on the Douyin platform

Distribution characteristics of city image short-video content

Descriptive statistics were performed on short-video information collected from the six cities, including the number of valid short-video samples, the length of the videos, the composition of the video makers and the level of interaction with users. The statistic results of the distribution characteristics and dissemination differences of short-video content among different cities were analyzed by mean, median and proportion and presented in Table 2.

Table 2. Descriptive statistics of sample data.

Statistical indicators	Beijing	Shanghai	Chengdu	Hangzhou	Xi'an	Chongqing	Overall
Valid samples (n)	486	479	493	482	488	491	2,919
Sample validity rate (%)	97.2	95.8	98.6	96.4	97.6	98.2	97.3
Mean video duration (s)	42.6	38.9	35.7	41.3	46.2	37.8	40.4
Duration standard deviation (s)	24.8	22.6	20.4	25.1	27.3	21.7	23.9
Government accounts (%)	16.5	14.4	11.8	15.1	13.5	12.2	13.9
News media accounts (%)	21.4	19.6	15.6	18.3	17.2	16.1	18.0
Commercial accounts (%)	18.9	24.0	25.8	22.6	21.5	24.6	22.9
Individual users (%)	43.2	42.0	46.8	44.0	47.8	47.1	45.2
Mean likes (n)	3,842	4,216	5,638	3,965	4,728	5,914	4,725
Median likes (n)	918	1,046	1,382	967	1,156	1,473	1,148
Mean comments (n)	186	204	278	193	231	296	232
Median comments (n)	39	46	61	42	52	67	51
Mean shares (n)	294	326	487	308	375	512	384
Median shares (n)	57	64	103	59	78	112	78

In total, the validity of the 2,919 valid samples were 97.3% with the mean duration of the video was 40.4 s, as shown in Table 2. Ordinary users made up for 45.2% of the total which was greater than other content creators. Chongqing and Chengdu were the only cities with mean like counts of 5,914 and 5,638, respectively. The median number of likes was 1,148 and the mean was 4,725, showing the interaction number of likes to be clearly right skewed. There were some variations in dissemination performance between cities and a handful of highly interactive videos moved the needle on the overall mean.

Analysis of city image themes and narrative characteristics

Based on 2,919 valid short-video samples collected from the Douyin platform, this paper conducted a categorical

analysis of city image themes and narrative styles; the results are shown in Figure 3.

Life experience videos accounted for 31.4%, historical and cultural videos for 24.6%, natural scenery videos for 19.8%, modern development videos for 13.7%, and social humanities videos for 10.5%, indicating that city images primarily focus on daily life and cultural resources. Such a distribution reveals that ordinary creators prefer to record trivial, authentic daily scenes close to audiences, rather than overly grand official propaganda content. In terms of narrative styles, experiential narratives accounted for 35.8%, expository narratives for 29.6%, and explanatory and story-based narratives for 21.3% and 13.3%, respectively.

Life experience themes predominantly employ experiential narratives. By contract, historical and cultural themes emphasize explanatory expressions. This reflects that distinct characteristics among different themes in terms of content organization and the presentation of urban symbols.

This matching relationship between theme types and narrative modes suggests that creators consciously select suitable storytelling strategies to fit their content orientation. Experiential narratives can bring immersive perception for viewers, whereas explanatory narratives help sort out context and deliver systematic cultural information. They construct a diversified and multi-layered urban media image on the short-video platform.

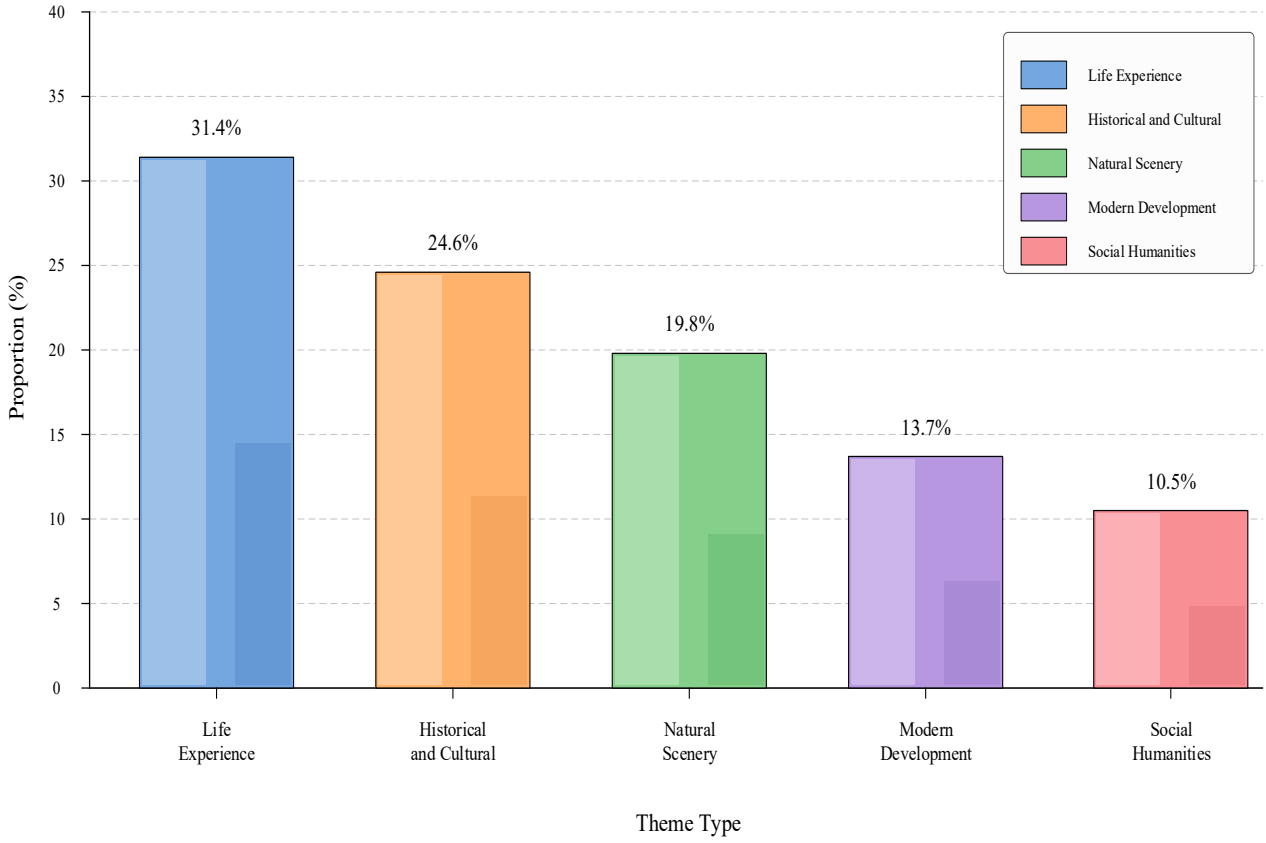


Figure 3. Distribution of urban image themes.

Analysis of visual symbols and emotional expression characteristics

It conducted a multi-label analysis on the 2,919 valid samples to examine differences in emotional expressions of various visual symbols. These visual symbols cover urban landmarks, architectural structures, neighborhood scenes, food, human figures and natural landscapes. An sentiment score range of -1 to 1 was defined and violin plots were used to illustrate the density, median and interquartile range of scores. The results are reported in Figure 4. Videos with food scored a median of 0.64, those with natural scenery scored a median of 0.59, both being higher than the median for building (0.38). The median scores for street and for people were 0.47 and 0.51, respectively.

The score distribution of cuisine was quite compact and people were more dispersed, indicating that there were differences among the different visual symbols concerning the intensity and stability of emotional expression.

The urban image is not only embodied in the picture plane but also produces different emotional expressions.

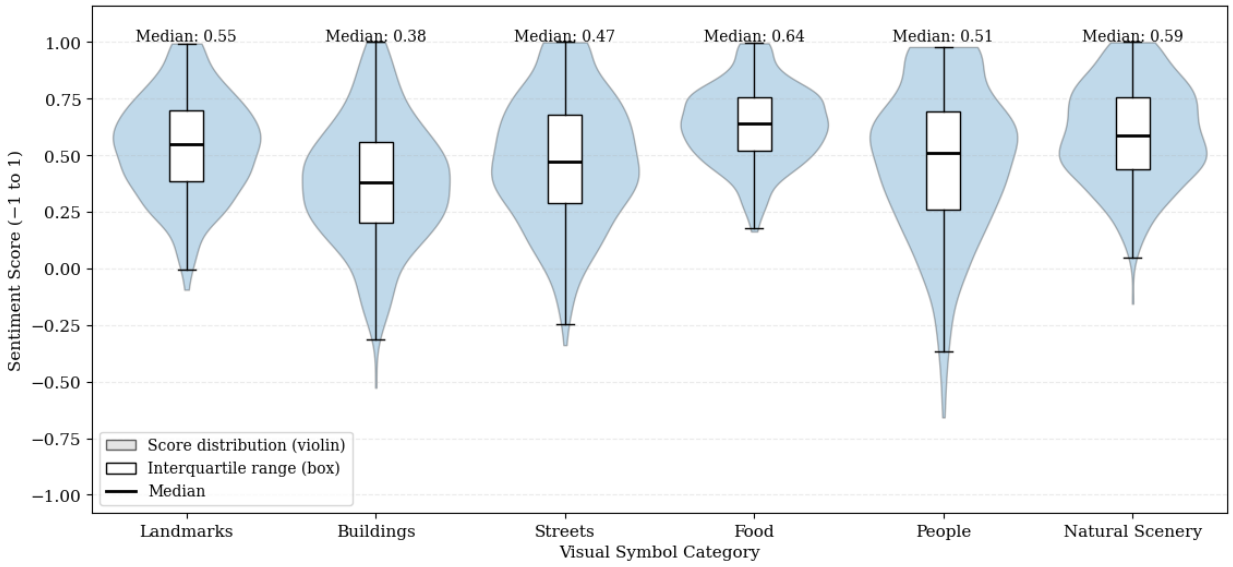


Figure 4. Correlation analysis of visual symbols of urban image and emotional orientation.

Differences in urban image construction among various communication actors

Differences in content selection and expression among different communication entities in the construction of urban image led to the classification of 2,919 short videos into four categories: government agencies, news media, commercial organizations, and individual users. A comparative analysis was conducted across dimensions such as thematic types, narrative styles, emotional tendencies, and user interaction levels to reveal their varied communication features. The specific results are shown in Table 3.

Table 3. Differences in city image content among diverse communication actors.

Statistical indicators	Government agencies	News media	Commercial organizations	Individual users
Sample size (n)	406	525	669	1,319
Historical and cultural themes (%)	28.8	27.6	22.4	23.6
Life experience themes (%)	14.5	20.8	39.6	35.9
Explanatory narratives (%)	38.7	35.2	16.4	13.8
Experiential narratives (%)	12.3	17.9	42.6	43.5
Positive sentiment (%)	78.6	68.4	82.5	73.7
Mean likes (n)	3,256	4,179	5,038	5,146
Mean comments (n)	142	207	238	263

Ordinary users accounted for 45.2% of the sample, with experiential narratives comprising 43.5% and an average of 5,146 likes, reflecting a strong emphasis on everyday life. Explanatory narratives from government agencies accounted for 38.7%, highlighting their role in information dissemination.

Lifestyle-themed content from commercial entities accounted for 39.6%, with positive emotions comprising 82.5%. News media content on historical and cultural themes accounted for 27.6%. Different entities exhibited distinct characteristics in content selection and narrative strategies, collectively forming a diverse system for expressing the city's image.

Analysis of the relationship between urban image content characteristics and user interaction

To explore the relationship between urban image content characteristics and user interaction behavior, 2,919 short videos were analyzed. Content themes, narrative styles, visual symbols, and emotional tendencies were used as explanatory variables, while the number of likes, comments, and shares served as interaction metrics. A negative binomial regression model was employed for

the analysis, and the results are shown in Figure 5. The "likes" incidence rate ratio (IRR) for lifestyle experience content was 1.32, for experiential narratives 1.26, and for positive emotional expressions 1.21 - all higher than the corresponding reference categories. The share incidence ratio for historical and cultural themes was 1.18, indicating a positive correlation with sharing behavior. In contrast, the comment incidence ratio for explanatory narratives was 1.03, indicating relatively weak interaction. Different content characteristics exhibit varying associations with user interaction levels. Everyday narratives and positive emotional expressions may be key factors influencing interaction in the dissemination of urban images.

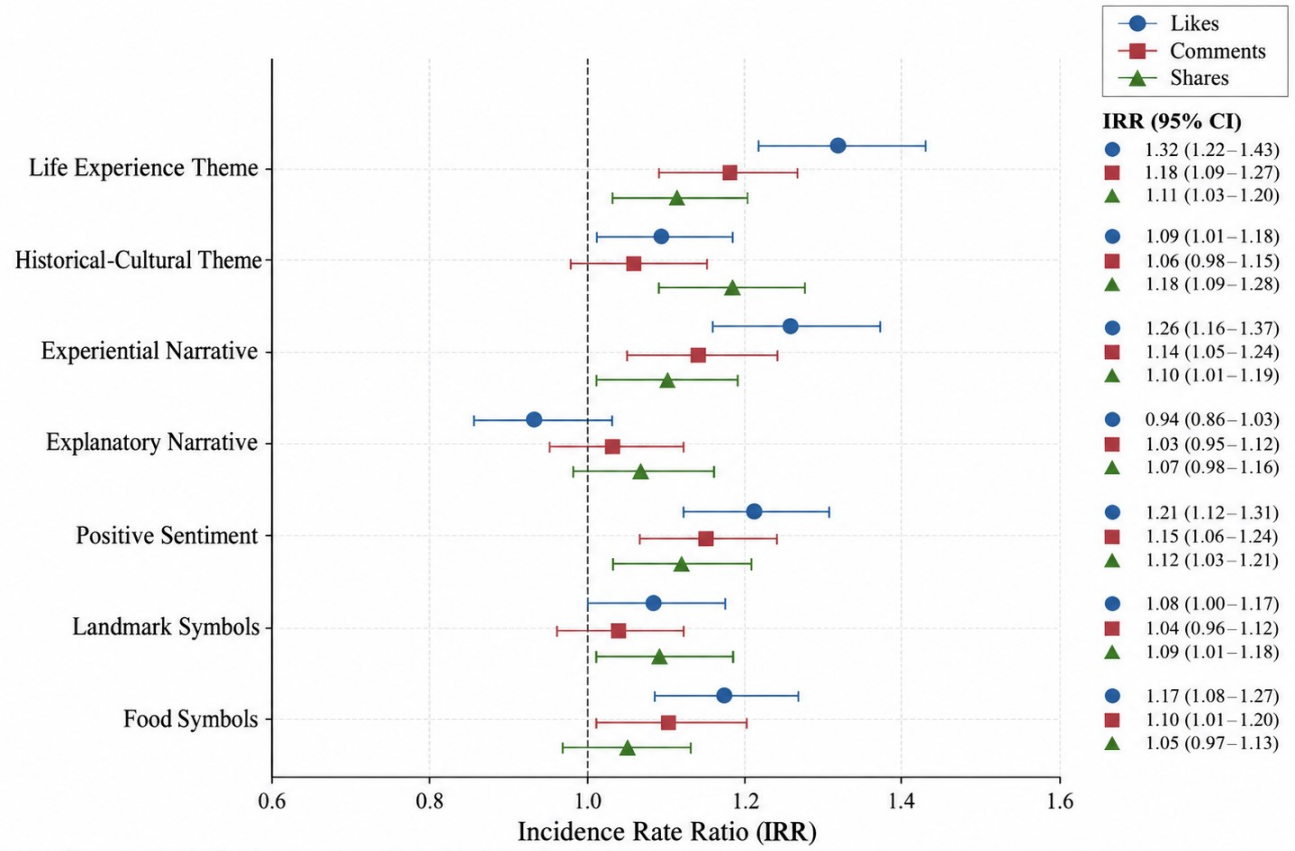


Figure 5. Relationship between city image content characteristics and user interaction.

Mechanisms and influencing factors in city image construction on the Douyin platform

City image construction on the Douyin platform is a complex process that relies on various factors, such as content selection, symbolic representation, emotional storytelling, and interactive feedback. Ordinary users (45.2%) and lifestyle experience themes (31.4%) suggest that the involvement of a wide range of actors and the expression of everyday life content is important in presenting image of the city. Visual symbols vividly depict urban spaces, such as city symbols, significant buildings, local food culture and human activities, and emotional narratives further emphasize the perceptual attributes of urban culture and lifestyle scenes.

The results of the interactive correlation analysis show that the like rates of lifestyle experience themes, experiential narratives, and positive emotional expressions are 1.32, 1.26, and 1.21, respectively. This finding demonstrates that content features have varied interactive associations with users.

Content creators' attributes, content themes, narrative strategies and visual elements jointly shape the dynamic process of urban image construction. The function of these factors is moderated by platform recommendation mechanisms and user feedback. These feedback signals further affect the visibility of content dissemination.

Conclusion

The construction of city image on the Douyin platform has characteristics of communication, with multi-actor collaborative participation, multimodal symbolic expression, emotional interaction. How city images are constructed in the media through the selection of topic, the encoding of visual symbols and the differentiation of narration gives it local identity. Urban image can be effectively conveyed through lifestyle experience

content and historical and cultural materials. Meanwhile, communication subjects differ noticeably in their selection of themes, narrative arrangements and emotional expressions.

This study adopted a multidimensional content coding system and conducted mixed analyses of textual, visual and emotional characteristics. It expands the quantitative research pathways for urban image construction and reveals the mechanism linking content characteristics to user interaction. Some limitations in the study exist, however, in terms of the sample cities and the period of study, as well as the lack of direct verification of the mechanisms of platform recommendation and changes in user cognition. For future research directions, scholars may integrate cross-platform comparison, longitudinal observation and audience perception surveys. This approach can further advance empirical inquiries into the dynamic mechanisms of urban image construction and its communication effects.

Funding

This research was not supported by any funds.

Acknowledgements

The author would like to show sincere thanks to those technicians who have contributed to this research.

Conflicts of Interest

The author declares no conflict of interest.

References

[1] Li, X. (2025) Manufacturing the influencity: short-form video and the politics of visibility in China's new first-tier cities. *Geoforum*, 167, 104481.

[2] Wang, F., COSTA, F. D., SABRI, M. F. (2025) TikTok short video platform and the construction and dissemination of urban image. *Quantum Journal of Social Sciences and Humanities*, 6(5), 413-423.

[3] Wang, F., COSTA, F. D., SABRI, M. F. (2025) The impact of TikTok user-generated content on the urban image of Tianjin, China. *Quantum Journal of Social Sciences and Humanities*, 6(5), 286-297.

[4] Zhuang, X., Wang, Y. (2025) Study on the construction and communication strategy of urban tourism image from the perspective of mobile short videos. *Lecture Notes in Education, Arts, Management and Social Science*, 3(8), 305-310.

[5] Liu, N., Sun, X., Hong, S., Zhang, B. (2022) Reproduction, cultural symbolism, and online relationship: Constructing city spatial imagery on TikTok. *Frontiers in Psychology*, 13, 1080090.

[6] Song, X. (2025) A study on the dual effects of regional image construction in Chinese short videos: a case study of tourism enthusiasm in Northeast China. *Global Dialogue on Media Dynamics, Trends and Perspectives on Public Relations and Communication*, 364-370.

[7] Zhang, W., Chen, J., Ni, Z., Si, X. (2026) Short videos and destination image reconstruction: a multidimensional comparative analysis on three Chinese wanghong tourism cities. *Asia Pacific Journal of Tourism Research*, 1-20.

[8] Zhao, J., Zhang, D. (2024) Visual propaganda in Chinese central and local news agencies: a Douyin case study. *Humanities and Social Sciences Communications*, 11(1), 588.

[9] Cachulo Pereira, M., Mendes Bacalhau, L. (2026) Scroll, watch, travel: an integrative framework of short-form video influence in destination marketing and place branding. *Place Branding and Public Diplomacy*, 1-23.

[10] Lin, H., Wen, H., Ma, Z. (2024) A study on the construction of destination image for China's county-level integrated media centers: a case study of four counties in Fuzhou. *Frontiers in Communication*, 9, 1346212.

[11] Ahmadpour, S., Wan, T. C., Toghrayee, Z., HematiGazafi, F. (2017) Statistical analysis of video frame size distribution originating from scalable video codec (SVC). *Complexity*, 2017(1), 8098574.

[12] Yan, P., Yu, C., Zhang, C., Shi, H., Wang, S. (2026) From pixels to perceptions: analyzing rural tourism image diffusion through Douyin-style short videos. *Asia Pacific Journal of Marketing and Logistics*, 1-17.

[13] Liu, Z. (2025) A Study on the Audience of Cantonese Short-Form Video Content from the Perspective of Place Identity Theory. *Journal of Modern Social Sciences*, *2*(4), 269-280.

[14] 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.

[15] Meng, J., Zhao, H. (2023) Report on the status quo and development of short-form video research in China. *Videolised Society*, 277-324.

[16] Li, K., Rui, L. (2025) The impact of short-form videos on tourist travel intention toward cities: an integrated theory of planned behavior and elaboration likelihood model approach. *Humanities and Social Sciences Communications*, 12(1), 1997.

[17] Zhang, C., Zheng, H., Wang, Q. (2022) Driving factors and moderating effects behind citizen engagement with mobile short-form videos. *IEEE Access*, 10, 40999-41009.

[18] Zhu, H., Wei, H., Wei, J. (2024) Understanding users' information dissemination behaviors on Douyin, a short video mobile application in China. *Multimedia Tools and Applications*, 83(20), 58225-58243.

[19] Wei, T., Wang, X. (2022) A historical review and theoretical mapping on short video studies 2005-2021. *Online Media and Global Communication*, 1(2), 247-286.

[20] Paquienseguy, F., Guo, Q. (2025) Douyin and the digital spread of intangible cultural heritage: Transforming cultural dissemination in the short videos age. *Emerging Media*, 3(2), 343-366.

[21] Zou, S. (2023) Curating a scopic contact zone: short video, rural performativity, and the mediatization of socio-spatial order in China. *Television and New Media*, 24(4), 452-470.

[22] Tang, H., Liu, J., Peng, X., Yang, Y., Liu, X. (2026) Word-of-mouth marketing in the digital age: Leveraging Short-video platforms for rural tourism marketing. *Journal of Theoretical and Applied Electronic Commerce Research*, 21(9), 286.