Education Reform and Development, 2026, Volume 8, Issue 10
https://bbwpublisher.com/journal/erd
Online ISSN: 2652-5372
Print ISSN: 2652-5364

A Study on the Aesthetic Discourse Construction Path of Cultural Tourism Media Short Videos for City Image Promotion from the Perspective of Appraisal Theory and Appreciation Resources

Linshang Li*

Nanfang College Guangzhou, Guangzhou 510970, Guangdong, China

Author to whom correspondence should be addressed.

Copyright: © 2026 Author(s). This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY 4.0), permitting distribution and reproduction in any medium, provided the original work is cited.

Abstract: Against the background of deep integration of cultural-tourism development and short-video-embedded urban communication, the construction of aesthetic discourse in city-image-promotion short-videos has become an issue calling for urgent attention. Taking appreciation resources within the Appraisal Theory of Systemic Functional Linguistics as the analytical perspective, this paper focuses on city-image-promotion short-videos released by Guangzhou cultural-tourism media and investigates the mechanisms and dilemmas in their aesthetic-discourse construction. The study finds that in the allocation of three categories of appreciation resources: reaction, composition and valuation, current short-videos have problems such as spectacle aesthetics overshadowing everyday aesthetics, lack of rhythm in image combination, labelling of value symbols, and cross-modal coordination fracture, which restrict the depth and recognizability of aesthetic expression of Guangzhou's city image. Based on typological analysis of appreciation resources, this paper refines four construction mechanisms: qualitative transposition, rhythmic allocation, contextual anchoring and attitudinal prosody coordination. It further puts forward optimization strategies, including double-layer allocation of "spectacle-everyday", rhythmic image narration, re-contextualization of cultural symbols, and cultivation of audiences as appreciation subjects, and proposes concrete improvement suggestions, taking Guangzhou as an example. This research expands the application of Appraisal Theory in multimodal cultural-tourism discourse and provides linguistic references for city-image communication.

Keywords: Appraisal theory; Appreciation resources; Cultural-tourism short-videos; City image; Aesthetic discourse; Guangzhou

Online publication: November 25, 2026

1. Introduction

Short videos have become a mainstream form of city-image communication. As a core city of the Guangdong-Hong-Hong-Kong-Macao Greater Bay Area and a centre of Lingnan culture, Guangzhou's cultural-tourism

media has been releasing city-image-promotion short-videos on platforms such as Douyin, WeChat Channels and Xiaohongshu, for example, Welcome to Guangzhou, Blooming Guangzhou, and various works themed on "Twelve Hours in Guangzhou". These short-videos not only perform information-notification functions, but also constitute aesthetic practices, constructing aesthetic experiences of Guangzhou jointly through images, sounds and texts.

Nevertheless, obvious tensions exist in the current urban aesthetic discourse of Guangzhou cultural-tourism short-videos. On the one hand, drone aerial photography and time-lapse photography render landscapes such as the Canton Tower, Zhujiang New Town and Baiyun Mountain into highly visualized presentations. On the other hand, a large number of works fall into formulas of high-altitude overlooks, landmark buildings, and lyrical background music. Aesthetic evaluations remain confined to reactive words such as "stunning", "shocking" and "prosperous", lacking in-depth appreciation of urban fabric, daily life and the values of Lingnan culture. Appraisal Theory from Systemic Functional Linguistics provides a refined analytical tool for investigating this phenomenon. Appreciation resources within its attitude subsystem are specially designed to describe speakers' or producers' aesthetic evaluations of things, processes and phenomena, covering three dimensions: reaction, composition and valuation. Extending appreciation resources from verbal texts to multimodal short-video discourse enables observation of the shaping processes, faults and improvement approaches of urban-image aesthetic discourse.

This paper mainly studies city-image-promotion short-videos released by Guangzhou cultural-tourism media. From the perspective of appreciation resources in Appraisal Theory, it analyses dilemmas in the construction of their aesthetic discourse, and further puts forward implementable construction paths, so as to offer references for linguistic research on city-image communication and practices of cultural-tourism media.

2. Theoretical overview

2.1. Appraisal theory and appreciation resources

Systemic Functional Linguistics regards language as a social-semiotic system, studying how language simultaneously fulfils three metafunctions: ideational, interpersonal and textual. Based on the interpersonal metafunction, Martin et al. put forward Appraisal Theory, which systematically describes meaning resources deployed by speakers or writers to express attitude, engagement and graduation. The attitude system consists of three subsystems: affect, judgement and appreciation. Affect refers to emotional reactions; judgement denotes ethical evaluations of behaviours; appreciation stands for aesthetic evaluations of objects, products, processes and phenomena [1]. Appreciation resources fall into three categories: reaction, which concerns whether something is appealing or pleasant, encompassing impact and quality; composition, which concerns whether something is balanced, complex and coherent; valuation, which concerns whether something bears social significance, innovativeness and importance.

Appreciation resources are not scattered vocabulary items, but meaning potentials that develop from explicit to implicit and from core to peripheral. In discourse, appreciation resources are realized by adjectives, nouns, verbs, clauses and rhetorical structures. They are not merely linguistic phenomena, but can also be manifested via multi-modal semiotic resources such as vision and audition [2]. In visual discourse, colour saturation, compositional balance and shot rhythm can correspond to specific resource elements; in auditory discourse, musical styles and ambient sounds also serve as specific resources to arouse aesthetic reactions.

2.2. Aesthetic-discourse construction of short-videos

The aesthetic-discourse construction of short-videos refers to the constructive process in which producers select, combine and evaluate urban landscapes, life scenarios and cultural images by means of semiotic resources including shots, frames, sounds, language and subtitles, so as to form aesthetic experiences and meanings of cities. Different from traditional verbal discourse, the aesthetic discourse of short-videos is multimodal, chronological and fragmented [3]. It must capture audience attention, trigger emotional resonance and leave memorable impressions within a very short timespan.

Multimodal Discourse Analysis in Systemic Functional Linguistics holds that visual symbols also possess representational, interactive and compositional functions; music, ambient sounds and voice-overs jointly constitute attitudinal prosody. Appreciation resources are generally distributed in short-videos following the pattern of "attitudinal prosody". For instance, a high-altitude aerial shot conveys the reactive meaning of "shocking" through colour and motion; slow-motion footage of street-market daily life delivers compositional meanings via frame balance and detail presentation; a publicity slogan such as "Here is Guangzhou, the millennium-old commercial metropolis" expresses valuative meanings through language. The allocation proportion, combination patterns and modal coordination of the three types of appreciation resources directly influence the thickness and depth of short-video aesthetic discourse. Over-accumulation of reactive resources will lead to aesthetic fatigue; absence of valuative resources deprives city images of cultural depth; deficiency of compositional resources renders frames either disorderly or insipid. Therefore, appreciation resources of Appraisal Theory can supply operable descriptive frameworks and diagnostic tools for the aesthetic-discourse construction of short-videos.

3. Dilemmas in aesthetic-discourse construction of cultural-tourism-media city-image-promotion short-videos

3.1. Excessive foregrounding of reactive appreciation resources: Spectacle aesthetics overshadowing everyday aesthetics

Guangzhou cultural-tourism-media city-image-promotion short-videos commonly narrow "beauty" down to "visual spectacle". In numerous short-video works, opening shots mostly feature drones flying over the Pearl River, with the Canton Tower, East Tower and West Tower rising amid lights, accompanied by voice-overs or subtitles such as "shocking", "resplendent" and "reaching new heights of beauty" [4]. Such reactive appreciation resources are intensively foregrounded, generating attention peaks within the first few seconds. The problem lies in the fact that reactive resources only focus on "impact", namely sensory stimulation and emotional arousal, rarely touching upon "quality", let alone composition and valuation. As a result, Guangzhou's urban aesthetic experience is reduced to the visibility of night-scene lighting and skyscraper clusters.

Everyday-aesthetic objects with delicate textures such as granite-flagstone streets in Yongqingfang, colonnades of Qilou buildings along Longjin Road, bodhi leaf shadows at Guangxiao Temple, and Manchurian-style windows in Xiguan Mansions usually flash past in quick-cut shots without shot dwelling or appreciative coding. Audiences only obtain a general impression that "Guangzhou is prosperous", yet cannot form a structural understanding of "why Guangzhou is beautiful".

3.2. Weak compositional appreciation resources: Lack of rhythm and hierarchy in urban-image combination

Compositional appreciation resources emphasize balance, complexity, and a sense of order of objects. In city-image short-videos, they are reflected in whether frame composition, shot rhythm, scene transition and symbol combination form aesthetic structures of meaning. A common defect in Guangzhou cultural-tourism short-videos lies in overly repetitive image selection and insufficient hierarchical combination [5]. Symbols including Pearl-River night cruises, morning-tea dim-sum, Cantonese-opera facial masks, lion-dance performances, flower fairs and Qilou buildings are juxtaposed in rapid succession without logical connections in time, space or narration.

Cuts from skyscraper clusters to morning-tea shrimp dumplings and then to lion dances produce individually well-composed frames, yet no progressive relations from modernity to locality or from public space to gustatory memory. Deficient compositional appreciation resources further result in even-paced shots devoid of breathing rhythm, spliced-instead-of-transitional scene shifts, and inconsistent colour styles. "Coexistence of the old and new, integration of mountains, rivers, cities and farmlands, and fusion of Chinese and Western elements", which are inherent complexities of Guangzhou's urban space, can generate rich contrasts and layers, yet are frequently simplified into uniform enumeration in short-videos.

3.3. Superficial valuative appreciation resources: Labelled consumption of regional cultural symbols

Valuative appreciation resources stress judgements on the social significance and cultural values of objects, constituting a vital link for city images to evolve from "visually appealing" to "meaning-rich". It is not that Guangzhou cultural-tourism short-videos contain no valuative discourse, but that such discourse is often superficial and slogan-oriented. Voice-overs frequently deploy labels such as "millennium-old commercial metropolis", "centre of Lingnan culture" and "Eating in Guangzhou", yet seldom interpret their cultural logics through concrete frames and details [6].

For example, when Qilou buildings are adopted as Lingnan-architectural symbols, shots mostly focus on facades and rarely enter colonnade spaces, failing to demonstrate their social functions of sun-shading, rain-sheltering, commercial-residential integration and continuation of street-side life. When Cantonese opera is presented via facial-mask close-ups or stage clips, value dimensions such as vocal styles, role types and community inheritance are not elaborated. Consequently, valuative appreciation resources degenerate into hollow accumulation of nouns, rather than judgements on the intrinsic qualities and social-context meanings of objects.

3.4. Insufficient coordination of cross-modal appreciation resources: Mutual undermining of audiovisual and verbal discourse

The aesthetic discourse of short-videos is multimodal, composed of visual, auditory, and verbal texts. Only when appreciation resources realize cross-modal synergistic effects can the outcome exceed simple addition of individual modalities. At present, Guangzhou cultural-tourism-media short-videos suffer from inadequate cross-modal coordination, which manifests in three specific forms.

First, vision and voice-overs are disconnected. For instance, frames display the sober order of white-collar commuters in Zhujiang New Town, while voice-overs adopt reactive phrases such as "passionate Guangzhou" and "vibrant flower city" to express affection for the city. Inconsistent attitudes between language and frames

lead to ambiguous aesthetic meanings. Second, mismatches between background-music rhythms and footage rhythms: slow-life street-market footage is paired with soaring electronic music, while fast-paced modern-city footage is matched with soothing guzheng melodies. Reactive meanings generated by music counteract compositional meanings conveyed by visuals. Third, excessive subtitle interpretation: numerous short-videos overlay large-character words such as “beautiful”, “shocking” and “inheritance” onto frames in an attempt to impose attitudes directly, which in turn inhibits audiences’ independent aesthetic appreciation [7]. Due to deficient cross-modal coordination, appreciation resources can hardly form continuous attitudinal prosody. The aesthetic discourse of the city image makes sense locally yet appears fragmented and hollow as a whole.

4. Construction mechanisms of short-video aesthetic discourse from the perspective of appreciation-theory appreciation resources

4.1. Qualitative transposition mechanism for reactive appreciation resources

Reactive appreciation resources shall not be confined to evoking superficial impact-dimension feelings such as “shocking” and “beautiful”. Their in-depth potential lies in transforming “impact” into “quality”. Impact refers to instant stimuli exerted by objects on perceivers, while quality relates to identifiable aesthetic attributes inherent in objects [8]. In constructing short-video aesthetic discourse, effective deployment of reactive resources requires shots to shift from “visibility of landscapes” to “readability of textures”.

In the case of Guangzhou, the “resplendence” of Pearl-River night views represents impact, whereas the “mildness” of washed-stone walls under Qilou colonnades in morning light embodies quality. The core of this construction mechanism lies in converting reactive resources from stimulus-response type to quality-perception type by means of shot parameters and temporal duration. Macro shots capture the translucent textures of coloured glass on Manchurian-style windows; slow panning shots render the sheen of rain-soaked bluestone pavements; static long shots depict steam rising from tea-house steamers against backlight. Such techniques rely on the textures of objects themselves rather than grand visual shocks, endowing reactive appreciation resources with qualitative depth. Qualitative transposition is also reflected in lexical selection: hollow adjectives are replaced by perception-oriented expressions such as “crystal-clear”, “mild”, “down-to-earth vibe” and “aged yet intact”.

4.2. Rhythmic allocation mechanism for compositional appreciation resources

Compositional appreciation resources emphasized balance, complexity, and a sense of order. Meanings of short-videos are mainly generated through shot-duration allocation, shot-size transition, scene connection and colour matching. The rhythmic-allocation mechanism dictates that compositional resources should not be spread evenly; instead, aesthetic rhythms of “exposition-development-climax-resolution” shall be established according to semantic layers of urban images [9].

Taking Guangzhou as an example, urban-space features such as “coexistence of old and new, coexistence of fast and slow, and interweaving of mountains, rivers, cities and farmlands” provide abundant inherent compositional materials. This construction mechanism requires producers to set contrasts and echoes in montage editing. For example, vertical-rhythm footage of Zhujiang New Town can be followed by transitional shots of subway rides or pedestrian movement, which then lead to horizontal extensions of Xiguan Qilou buildings; alternatively, natural negative-space footage of Baiyun Mountain can serve as a starting point, followed by the commercial density of Beijing Road. Shot durations shall maintain breathing rhythm:

landmark landscapes can adopt quick cuts for impact, while everyday scenarios require appropriate dwelling time for audiences to form spatial cognition. Uniform filter styles are not mandatory for colour schemes; contrasts between cool-toned glass curtain-walls in new urban districts and warm-toned grey brick walls in old urban districts may be deliberately retained. Structural "unevenness" echoes the complexity of Guangzhou's urban forms. The in-depth logic of rhythmic allocation stresses that compositional appreciation resources refer not merely to intra-frame balance, but also to order design across shot sequences.

4.3. Contextual anchoring mechanism for valuative appreciation resources

Valuative appreciation resources cannot be detached from concrete contexts and reduced to labels. The contextual-anchoring mechanism highlights that value judgements shall be attached to perceivable spaces, characters or processes. Cultural values are excavated from concrete scenarios via narrative or scenario-oriented approaches.

Guangzhou possesses abundant valuative resources, including commercial traditions of the Maritime Silk Road, pioneering spirits of modern-era revolutions, experimental nature of reform-and-opening-up, and pragmatic aesthetics of Lingnan daily life. Nevertheless, such values shall not be merely announced by voice-overs in short-videos; value anchors shall be deployed. An effective anchoring approach follows the logic of "small entry point — profound context". For instance, start with hammer-striking sounds of a hand-coppersmith beside Liwan-Yong Canal. Based on fragmented footage of his working procedures, historical connections between Xiguan handicrafts and Guangzhou's identity as a commercial city can be derived. Taking the porch of an old building on Shamian Island as an anchor point, slow-pushing shots and detailed presentations reinforce the historical weight of Guangzhou as a modern-era Chinese window onto the outside world. The construction mechanism for valuative appreciation resources is essentially a technique of "meaning implantation". Abstract cultural judgements are transformed into observable spatial details, audible ambient sounds and perceivable human actions, grounding valuative discourse in experiential foundations [10]. Only in this way can valuative appreciation resources evolve from slogans to understanding, and establish the cultural depth of city images.

4.4. Attitudinal-prosody coordination mechanism for cross-modal appreciation resources

The multimodal nature of short-videos demands attitudinal-prosody rhythms formed among visuals, audios and verbal texts. Attitudinal prosody means that attitudinal meanings are not evenly distributed across discourse, but unfold with rises and falls, primary and secondary emphases. The core of the coordination mechanism lies in modal division-of-labour and mutual reinforcement. Vision mainly presents compositional meanings and qualitative details; audition builds reactive atmospheres and emotional tones; language delivers value judgements and cultural-context hints. The three modalities perform their respective roles while complementing each other [11].

Taking Guangzhou-themed short-videos as an example, a feasible coordination strategy is as follows: when frames depict colonnade spaces under Qilou buildings, visuals demonstrate order and composition through colonnade rhythms and light-shadow segmentation; ambient sounds capture reverberations of passing motorbikes and fragments of Cantonese-language greetings from shop-keepers; subtitles or voice-overs deliver the value judgement that "colonnades constitute the second parlour for Guangzhou residents". The three modalities share identical attitudinal orientations yet undertake different appreciation dimensions, avoiding content duplication or meaning undermining. Only when cross-modal modalities complement one another

rather than simply pile up, can the aesthetic discourse of the city image generate unified and rich attitudinal experiences within a short time span.

5. Strategies for constructing aesthetic discourse of city-image-promotion short-videos released by cultural-tourism media

5.1. The "spectacle-everyday" double-layer allocation strategy for reactive appreciation resources

City-image-promotion short-videos generally face a core contradiction: spectacular footage can grab audience attention, yet excessive accumulation of spectacles triggers aesthetic fatigue and superficializes city images. From the perspective of appreciation resources in Appraisal Theory, spectacle aesthetics relies on impact-oriented reactions — the sensory shocks brought by aerial drone shooting, light shows and time-lapse photography. By contrast, aesthetic experiences in urban everyday life, including textures of streets, layers of light and shadow, timbre of sounds and rhythms of life scenarios, are borne by quality-oriented reactions [12]. Optimization can be carried out from the following three dimensions.

First, establish a resource-allocation framework of "spectacle anchoring — in-depth exploration of everyday life". The function of spectacular footage shall be confined to capturing attention and rapidly forming an overall city profile, and its duration shall not exceed one-fifth of the total length of the video. The remaining major part shall be devoted to presenting objects with in-depth qualities. This requires shots to descend from high altitude to ground level, shift from full shots to close-ups and details, and change from quick cuts to static dwelling, so that the material textures, grain, light-shadow effects and sounds of objects themselves become aesthetic objects.

Second, realize the qualitative transformation of the linguistic modality. The linguistic dimension of reactive appreciation resources is highly prone to the trap of empty impact-oriented words. For instance, words such as "shocking", "stunning", "dream-like" and "breathtaking" are overused in numerous city-promotion texts and have nearly been semantically exhausted. In response, quality-prioritized lexical criteria shall be formulated: every adjective must answer "what kind of beauty it is", instead of merely stating that "it is beautiful". For example, instead of "beautiful street scenes", use "rain-soaked flagstone roads glowing with mild luster"; instead of "shocking skyline", use "glass curtain-walls weave dusk into golden hues".

Third, embed qualitative-verification links into the production workflow. When reviewing short-video scripts, cultural-tourism media shall consciously classify reactive words into impact-oriented and quality-oriented categories, calculate the proportion of impact-oriented words and set an upper limit. When the density of impact-oriented words becomes excessively high, production teams shall return to the footage layer and seek alternative quality-oriented descriptions.

5.2. The "rhythmic image narration" strategy for compositional appreciation resources

Compositional appreciation resources focus on balance, complexity and order. In city-image short-videos, compositional meanings are not realized through the composition of individual frames, but through relations among shot sequences, namely contrast, echo, progression and circular recurrence between shots. At present, many city-promotion short-videos suffer from homogeneous collage in composition: shots feature uniform duration, similar shot sizes and random transitions, and urban images are treated as symbol lists whose order can be adjusted arbitrarily [13]. Against such a backdrop, it is necessary to convert urban images from "lists"

into “narratives”, and construct compositional appreciation resources via rhythmic image narration.

First, adopt contrasting structures rather than enumerative structures. The essential tensions of a city image lie in spatial and temporal contrasts, such as the old versus the new, fast versus slow, grandeur versus subtlety, nature versus man-made constructions. From the perspective of shot sequences, opposing elements shall be deliberately arranged to appear alternately and echo one another. For example, vertical-rhythm shots of dense high-rises shall be followed by horizontally-extended shots of old-town blocks; bustling commercial core districts shall be followed by tranquil life scenarios.

Second, pursue “shot-size breathing” instead of homogeneous shot sizes. Restricted by short-video duration, producers tend to adopt medium and medium-close shots to convey information rapidly. Nevertheless, the depth of compositional appreciation requires a sense of breathing among different shot sizes. Long shots illustrate relations between cities and nature; medium shots demonstrate spatial functions and activities; close-ups capture material textures and details of daily life.

Third, take form matching as transition logic. The in-depth effect of compositional appreciation resources derives from formal echo relations between frames, including curve echo, texture echo and perspective echo. When a curved skyline appears in one shot and is followed by curved roof ridges of traditional buildings in another shot, audiences will subconsciously perceive the inherent continuity of urban forms.

5.3. The “contextual anchoring” strategy for valuative appreciation resources

Valuative appreciation resources refer to the social significance and cultural importance of objects. A common problem for city-image short-videos on the value dimension is labelling: a large number of cultural symbols are presented in the form of names and slogans, without providing audiences with contexts for understanding their value connotations [14]. Accordingly, supported by contextual anchoring, abstract cultural-value judgements shall be fixed onto concrete, perceivable spaces, processes and characters.

First, adopt process presentation instead of result demonstration. The value of urban cultural symbols lies not in the finished appearance, but in the social logic embedded within their operating processes. For instance, the value of traditional handicrafts does not reside in static finished products, but in hand movements, tool traces and material transformations; the value of food culture is not reflected in plated finished dishes, but in heat control, rhythms and bodily craftsmanship during production. This requires production teams to enter the “production process” of symbols and adopt process-oriented shots to show how cultural values are generated.

Second, adopt “spatial context” instead of symbol close-ups. The value of a historical building manifests not only in its facade form, but also in its relations with surrounding environments, as well as its functions and usage status in contemporary urban life. Therefore, avoid showing symbols in isolation; pull the camera back to present the spatial context where symbols are situated, namely how they are used, who uses them, and which old and new buildings they coexist with.

Finally, enrich value-understanding perspectives via historical depth. The depth of urban-cultural values takes shape over time, yet the temporal constraints of short-videos make the historical dimension most likely to be neglected. Techniques such as dissolves, split screens, insertion of old photographs and integration of historical video materials can be adopted to deliver visual hints of past-and-present contrasts within a very short time, endowing contemporary landscapes with historical depth.

5.4. "Attitudinal-prosody coordination" of cross-modal appreciation resources and audience-engagement strategy

Multiple modalities, including vision, audition, and language, jointly construct the aesthetic discourse of short-videos. However, attitudinal conflicts or mutual counteraction may occur among modalities. Improvements can be achieved through two strategies: "attitudinal-prosody coordination" at the production end and "cultivating audiences as appreciation subjects" at the communication end.

Attitudinal-prosody coordination at the production end requires defining the role of each modality in expressing appreciation resources. Specifically, the chronological pattern of "vision first, audition for continuation, language for thematic highlighting" can be adopted: vision first builds concrete perceptions; audition then extends emotional atmospheres; concise language highlights values at critical moments. After language output, reserve blank space for vision and audition to keep functioning [15].

Cultivating audiences as appreciation subjects at the communication end is a frequently-overlooked component of construction strategies. The aesthetic discourse of city images shall not be a closed text produced solely by creators; instead, it shall become a social process continuously generated with audience participation. To be specific, it covers three aspects. First, set open-ended aesthetic topics and invite audiences to answer aesthetic questions raised in official videos via texts, images or short-videos. Second, regularly release creation notes or behind-the-scenes analyses with appreciation-oriented guidance, explaining intentions for composition, colour and sound design to audiences without ruining aesthetic experiences, so as to gradually enhance public aesthetic-appreciation capabilities. Third, incorporate high-quality user-generated content into official communication sequences to form an interactive cycle of aesthetic discourse.

6. Suggestions on constructing aesthetic discourse for cultural-tourism-media city-image-promotion short-videos: Taking Guangzhou as an example

6.1. Suggestions on aesthetic allocation of spectacle and everyday life: Implementation of Guangzhou's "double-layer shooting list" and quality-oriented copywriting

Cultural-tourism media in Guangzhou can implement the "double-layer shooting list" system in short-video production.

The first layer consists of anchoring shots: no more than two Guangzhou landmarks per video, with duration controlled within 15 % of total runtime, captured via aerial shooting or time-lapse, mainly functioning to grab attention at the opening. The second layer covers in-depth-exploration-of-every-life shots, accounting for more than 60 % of total length. It is recommended to select resources from the "Guangzhou Everyday-Aesthetic Map". For example, start with the 6 a.m. Longjin Road morning market in the early morning, showing steam curling from steaming baskets under morning light, with the voice-over "Guangzhou wakes up with a wisp of steam" as an introduction. In the morning, focus on Panxi Restaurant, presenting light reflected on floor tiles as dim-sum carts roll by, with Cantonese morning-greeting murmurs as ambient sound. In the afternoon, shift to Guangxiao Temple, showing light-and-shadow gradations of bodhi-leaf shadows falling onto blue-brick floors with a fixed-position shot lasting more than eight seconds. In the evening, depict West Binjiang Road, using slow panning shots to show relations between banyan aerial roots and ripples on the Pearl River. Late at night, present Baoye Road food-stalls with close-up shots of chefs tossing woks amid rising cooking-steam.

For copywriting, implement the system of restricting impact-oriented words and prioritizing quality-oriented words. Apply qualitative verification in review procedures: every voice-over sentence describing urban beauty shall answer "what kind of beauty it is". For instance, replace "Guangzhou's night view is shocking" with "River breeze

shatters lamplight; Guangzhou's night flows like glazed glaze"; replace "Qilou buildings are beautiful" with "The weathered surfaces of washed-stone walls are handicrafts carved by time".

6.2. Guangzhou-oriented proposal for compositional rhythm: "Vertical-horizontal-vertical" template and Lingnan-form transition index

Cultural-tourism media in Guangzhou shall formulate the "Guangzhou Urban-Image Rhythm Template", with a closed-loop rhythm of "vertical-horizontal-vertical" as its core. For example, open with a three-second low-angle vertical shot of the Canton Tower, followed by two-second transitional shots of shaking footage inside a metro carriage or cycling across Haizhu Bridge, then a four-second medium shot showing horizontal extension of Qilou-building streets, next a three-second macro close-up of a coppersmith's hammering surface in Xiguan or pleats of shrimp dumplings, and finally end with a drone ascending back to a vertical perspective. This template transforms compositional resources from homogeneous collage into breathing-rich image sequences.

Moreover, formulate a Guangzhou-form-transition index for editing matching. For example, match curved glass surfaces of the Canton Tower with curved gable roofs of Xiguan Mansions; match Pearl-River ripples with glazeware reflections of Guangcai porcelain; match repeated perspective of Qilou colonnades with repeated colonnade perspectives of the Sun Yat-sen Memorial Hall; match cloud-and-mist contours of Baiyun Mountain with crown outlines of old trees on Shamian Island. Abandon uniform filters for full-length videos and preserve colour settings by zones. For example, adopt cool-grey-blue tones for Tianhe District, warm-grey-yellow tones for Liwan and Yuexiu Districts, and teal-green tones for Pazhou and areas along the Pearl River. Colour contrasts themselves shall embody Guangzhou's urban composition of "coexistence of the old and the new".

6.3. Guangzhou practice for re-contextualizing Lingnan cultural symbols: Three-in-one shooting guidelines

Cultural-tourism media in Guangzhou shall formulate three-in-one shooting guidelines integrating "process, space and history" for key Guangzhou cultural symbols. For Qilou-building shooting, start from the process dimension: capturing early-morning store-opening, goods-restocking and neighbourly chats under colonnades. On the spatial dimension, adopt low-angle shots entering colonnades to present colonnade depth, ground shadows and permeable relations between shop interiors and exteriors. On the historical dimension, apply dissolve transitions or split-screen techniques to show contrasts between the same Qilou-building streets in the 1930s and the present, with only one thematic voice-over remark on the historical layer: "The colonnades of Qilou buildings are a one-hundred-year gift Guangzhou offers to pedestrians".

To present the values of Cantonese opera, shoot at waterside stages of the Cantonese-Opera Art Museum and private opera gatherings on the third floor of Ronghualou. For instance, record young performers applying facial makeup backstage, master-and-apprentice guidance, and community rehearsals; demonstrate spatial relations between waterside stages and auditorium seats; adopt dissolve transitions inserting old photos of Hong Xiannü from the historical perspective.

For morning-tea shooting, select Panxi Restaurant or Lin Heung Tea House. Show rolling-cart speeds of dim-sum carts on the process dimension; reflect interactions between high tea-house ceilings and market noises on the spatial dimension; insert old-photo scenes of "one pot and two dim-sum servings" on the historical dimension. Specifically, build value-anchor archives for each symbol, noting shootable locations, accessible

time periods, cooperative personnel and available historical video materials, so that the re-contextualization strategy becomes a sustainable regular production mechanism.

6.4. Guangzhou implementation of multimodal coordination and audience cultivation: Sound-scape library, blank-space subtitles and interactive planning

Cultural-tourism authorities of Guangzhou shall take the lead in constructing the Guangzhou Sound-Scape Library. Systematically collect sounds including iron-wheel rumbles of morning-tea carts, Cantonese-chat ambient noises on Qilou-building streets, Pearl-River ferry whistles, market Cantonese price-quoting and steelyard-weight clinks, Dragon-Boat-Festival drumbeats, Spring-Festival flower-market bargaining voices, rainwater dripping down drain-pipes under old Qilou buildings, and teapot-whistling sounds in tea-houses. Classify and label sounds according to emotions and scenarios, providing local sound-scape alternatives to stereotyped commercial background music for editing.

Implement the "blank-space specification" for subtitles: no more than three thematic subtitles per video. Thematic subtitles shall deliver appreciation values on quality or valuation dimensions, and shall not repeat reactive meanings already conveyed by frames. For example, given footage showing Pearl-River ripples and joggers, the subtitle can read "The river lies within the city, and the city rests upon the river". Adopt low-density mid-bass voice-overs to embody the calm temperament in Guangzhou's urban aesthetics, supplemented by appropriate background music.

At the audience end, build an appreciation-interaction chain. Raise open-ended questions at the end of each episode, such as "What is the most beautiful sound of Guangzhou in your memory? Which Qilou-building street makes you willing to walk an extra hundred metres?" Invite users to respond and interact via texts or short-videos. Excellent user responses may be incorporated into subsequent official videos. Accordingly, the analytical potential of appreciation resources from Appraisal Theory can be extended from the production end to the communication end, enabling the continuous generation of Guangzhou's urban-image aesthetic discourse between producers and audiences.

7. Conclusion

To sum up, taking appreciation resources of Appraisal Theory within Systemic Functional Linguistics as the analytical framework, this paper focuses on the aesthetic-discourse construction of city-image-promotion short-videos released by Guangzhou cultural-tourism media. Theoretically, it expands the three-dimensional classification of appreciation resources — reaction, composition and valuation — from verbal texts to multimodal short-video discourse, turning them into analytical tools for describing and diagnosing urban-image aesthetic practices. Practically, this paper reveals four dilemmas existing in current Guangzhou cultural-tourism short-videos: excessive foregrounding of reactive resources leading to spectacles overshadowing everyday aesthetics; weak compositional resources resulting in rhythm-deficient image collage; superficial valuative resources causing labelled consumption of cultural symbols; insufficient cross-modal coordination giving rise to mutual counteraction between audiovisual language and verbal discourse. On such grounds, four construction mechanisms including qualitative transposition, rhythmic allocation, contextual anchoring and attitudinal-prosody coordination are refined. Discussions are carried out from both macro-strategic perspectives and Guangzhou-oriented implementation perspectives. Macro strategies apply to cultural-tourism short-video production for ordinary cities, while Guangzhou-specific

proposals are implemented into concrete shooting lists, rhythm templates, symbol-shooting guidelines and sound-scape-library construction.

Disclosure statement

The author declares no conflict of interest.

References

[1] Li QQ, Pu XF, 2025, Communication Characteristics, Development Dilemmas and Optimization Paths of Cultural-Tourism Short-Videos. Convergence Media, (9): 57–62.

[2] Huang Y, 2025, From Cultural Excavation to Emotional Resonance: Discourse Practice of City-Image Promo Datong · Non-Contention. Screen and World, 2025(12): 24–26.

[3] Zhu XT, 2025, Research on the Construction of Guangzhou's City Image by Government-Affiliated Cultural-Tourism Short-Videos From a Spatial Perspective, thesis, Guangdong Polytechnic Normal University.

[4] Peng ZH, 2025, Research on the Construction of Guangzhou's City Image by Morning-Tea Culture, thesis, Guangdong University of Foreign Studies.

[5] Hu YZ, 2025, Research on the Construction of China's National Image by Local Cultural-Tourism Short-Videos on TikTok, thesis, Guangdong University of Foreign Studies.

[6] Wang Q, 2025, Research on Short-Video Communication Strategies of Urban Media Image Under the Boom of "Student-Special-Forces-Style Tourism", thesis, Chengdu University.

[7 Lu YW, 2024, Research on Narrative Design of City Promotional Videos Based on Short-Video Platforms—Taking Time Capsule as an Example, thesis, Suzhou University of Science and Technology.

[8] Wang L, 2024, Cultural Identity, Empathetic Narration and Aesthetic Evolution: Multiple Dimensions of City-Image Construction Empowered by Short-Videos. New Media Research, 10(6): 63–65.

[9] Xie AQ, Zhang X, 2023, Image, Value and Reception: Visual-Rhetoric Construction in "Guo-Feng"-Themed Cultural-Tourism Short-Videos. Modern Audio-Visual Arts, 2023(10): 45–51.

[10] Wurina, 2023, Research on Guangzhou City-Image Promotional Videos, thesis, Chang'an University.

[11] He GP, Cai YY, 2023, Content Analysis of Global-Landscape-Oriented Guangzhou and Shenzhen City-Image Promotional Videos: 2010–2021. Southern Media Research, 2023(4): 60–70.

[12] Guo AQ, 2023, Research on Visual Communication of Guangzhou's City Image Based on Douyin Platform, thesis, Hebei University of Economics and Business.

[13] Liu C, Xiong KR, Zhang YX, et al., 2022, Urban-Symbol Construction and Image Shaping From the Perspective of Promotional Videos—A Content Analysis of 101 Guangzhou City-Image Promotional Videos From 2013 to 2020. Urban Observation, 2022(6): 112–127 + 163.

[14] Zhou WY, 2020, Research on the Communication of Guangzhou's City Image in Food-Themed Short-Videos, thesis, Guangdong University of Foreign Studies.

[15] Wang MY, 2020, Analysis of Communication Strategies of Guangzhou City-Image Promotional Videos From the Perspective of Semiotics. Audio-Visual Vision, 2020(2): 212–213.

Publisher's note

Bio-Byword Scientific Publishing remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.