

Image Structure of Cities Formed by Social-Networking Service Posts —Spatial Distribution and Content Similarity Evaluation on the Urban Landscape Images in Central Tokyo From Flickr

Tetsuya YAGUCHI^{1,*}, Takumi FUJINUMA²

¹ Department of Architecture, School of Creative Science and Engineering, Waseda University, Tokyo 169-8555, Japan

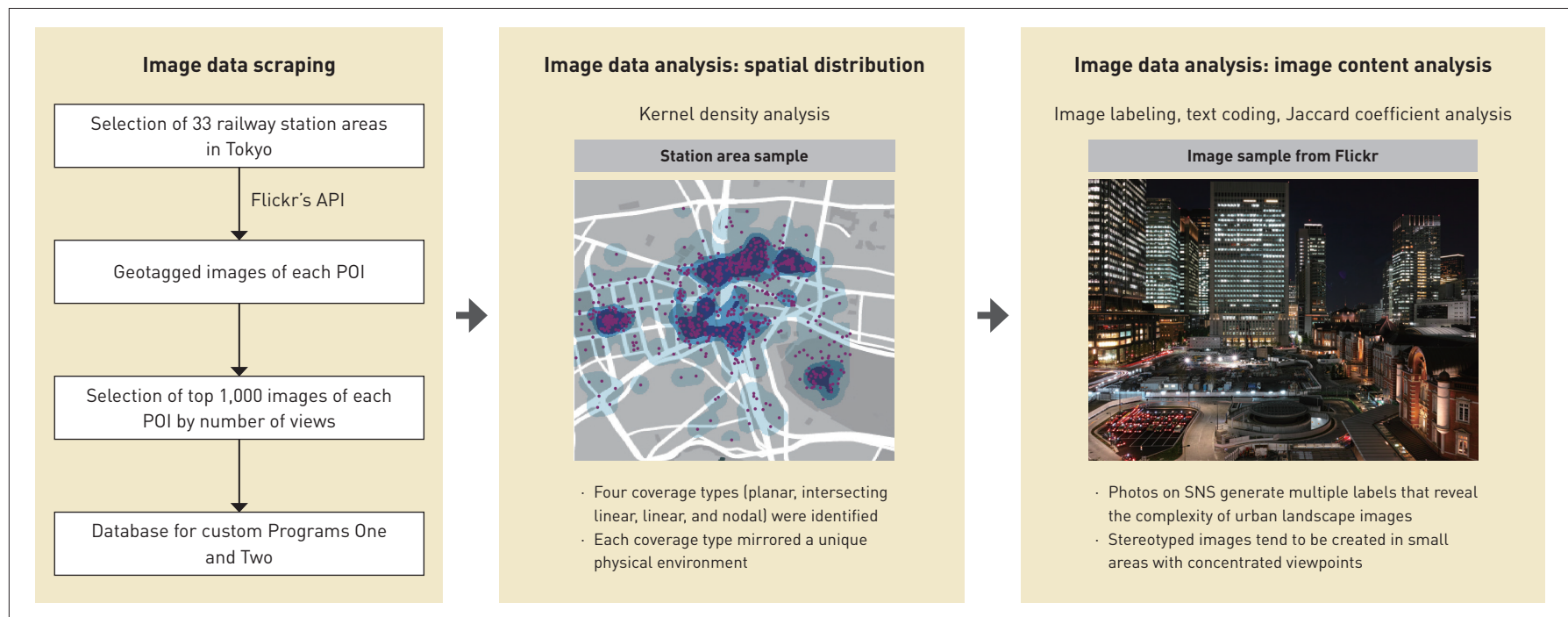
² Goodpatch, Inc., Tokyo 150-0032, Japan

*CORRESPONDING AUTHOR

Address: Okubo 3-4-1, Bldg. 55N, Ste. 704B, Shinjuku, Tokyo 169-8555, Japan

Email: tetsuya.yaguchi@waseda.jp

GRAPHICAL ABSTRACT



ABSTRACT

This research investigated the impact of social-networking service posts on the formation of image structure of cities, focusing on the spatial distribution of images and their content similarity. It aimed to delineate the image structure of cities created by numerous users, moving beyond traditional qualitative methods towards a more quantitative and objective approach with big data. Taking

central Tokyo as an example, this study extracted geotagged image data of 33 major railway station areas from Flickr's API (Application Programming Interface). Four coverage types of viewpoint distribution, namely planar, intersecting linear, linear, and nodal, were identified, reflecting the unique urban structures respectively. Further investigation of the image contents, primarily

consisting of “urban landscape” and “landscape/street trees,” showed that such contents significantly influenced the formation of the image structure of cities. The study concluded that as the number of photo posts increased and the representative viewpoints concentrated, the digital information received by users became more homogeneous, leading to strongly stereotyped images of urban landscapes. These findings highlight the role of social networking services in shaping perceptions of the urban environment and provide insights into the image structure of cities as formed by digital information.

KEYWORDS

Social Networking Service; Image Structure of the City; Flickr; Image Analysis; Stereotyped City Image; Perception; Tokyo

HIGHLIGHTS

- Social networking service posts can illustrate the image structure of cities using GIS mapping techniques
- The spatial distribution of viewpoints can be classified into planar, intersecting linear, linear, and nodal coverage types
- Increased uniformity of digital information leads to stereotyped perceptions of urban landscapes

EDITED BY Ying WANG, Xidong MA

1 Introduction

1.1 Research Background

In recent years, advancements in information technology and the widespread adoption of digital devices such as personal computers and smartphones have facilitated the exchange of information among individuals. Currently, numerous people, particularly the younger generations, use social media to share information, resulting in a vast accumulation of data on the Internet^[1]. Through

their extensive engagement with digital devices, users participate in the society by sharing experiences on social media and reacting to information shared by others. Such Internet usage enables users to receive considerable volume of information both passively and actively, with the digital content significantly influencing their perceptions of the urban environment^[2].

The prevalent use of smartphones has notably influenced people’s social behaviors and activities as well. In Japan, there has been a noticeable yearly decrease in population going outdoors^[3]. According to a survey of mobility^[4], individuals in their 20s, the smartphone-native generation, tend to go out less often than those in their 70s. Moreover, it has been observed that younger individuals, particularly those in their 20s and 30s, are inclined to document their experiences with photographs and share them on social networking services (SNS), while the proliferation of smartphones has spurred visits to previously unconsidered places^[4].

1.2 Research Objective

This study examined the position of recipients of digital information, especially photographs posted and archived on Flickr^[5]. By quantitatively analyzing urban landscapes, sceneries, and the subjects within these photographs, the study attempted to delineate the image structure of cities created through SNS posts by a wide user base. Previous research on urban images has largely employed qualitative methods such as mental and cognitive mapping^[6] and interview analysis, which often suffered from limited sample sizes and low objectivity. Utilizing the extensive data available on Flickr, this study introduced a novel, quantitative, and objective methodology.

1.3 Literature Review

1.3.1 Theoretical Framework of “the Image of the City”

In the field of urban planning, the concept of “the image of the city,” which encompasses the quality of individual’s perceptions and understandings of the urban environment, was first introduced by Kevin Lynch in 1960^[6]. He believed that people’s comprehension of urban spatial layout was a critical component of urban life. To gauge and illustrate the image of the city, Lynch proposed the notion of “imageability,” defining it as the quality in a physical object that heightens its chances of provoking a strong image in any given observer. This is attributed to the shape, color, or arrangement that eases the creation of vividly identified, powerfully structured, and highly useful images of the environment. Lynch identified five elements—namely paths, edges, districts, nodes, and landmarks—

that constitute the image structure of a city from an observer's viewpoint^[6]. Subsequent research over sixty years has affirmed these elements as stable components of the city image. For example, Donald Appleyard^[7] explored the multifaceted ways in which cities are perceived, designed, and structured, emphasizing the understanding of both the objective and subjective experiences of citizens. This perspective on urban perception and design was further echoed by Stephen Kaplan in the 1970s^[8] and Stephen Carr et al. in the 1980s^[9] with comprehensive reviews on Lynch's contributions. They underscored the importance of designing cities that are both easily navigable and relatable for their inhabitants, and concurred with Lynch's proposition that by fostering a clear image of the city, urban spaces can be rendered to be more meaningful and enriching for their residents. Jack L. Nasar, in his studies on Knoxville and Chattanooga of Tennessee in the USA, emphasized the significance of understanding how individuals perceive and evaluate urban environments^[10]. Through mapping evaluation, Nasar demonstrated that urban planners and designers could gain invaluable insights into areas that either resonate positively or negatively with residents and tourists. His findings align with Lynch's theoretical framework, reinforcing the idea that a clear city image is pivotal for residents' and tourists' positive experiences. From a branding perspective, Gert-Jan Hospers posited that Lynch's framework on the city's imageability can be instrumental for city marketers aiming to carve out a distinctive image for their cities^[11], indicating the impact of urban design principles on both resident experiences and city branding and marketing.

However, Lynch's methodologies have also been critiqued for potentially fostering a "stereotypical" city image due to their inherent oversimplification and subjectivity.^[12] Moreover, the traditional methods for image mapping, failing to capture the dynamic city nature, tend to offer a static perception for each of Lynch's urban elements. However, in reality, they are continuously evolving and interrelated, rather than independently existing.^{[13][14]}

1.3.2 The Relationship Between Photo Images on SNS and the Image Structure of Cities

In response to these identified shortcomings, Gabriele Filomena et al. developed a computational approach to spatial recognition, employing GIS-based technologies and space syntax theory^[14]. Liu Liu et al. discussed the potential of using geotagged images posted on social media as alternatives to traditional social surveys and interviews, allowing for the study of numerous data inputs^[15]. Shinichi Okuyama et al. suggested that city images are formed not

only through direct experiences, but also via media channels such as magazines and the Internet^[16], by demonstrating the influence of photo distribution and content on Google Maps in shaping Tokyo's image. Yuji Osaki et al.^[17] and Hiro Okutsu et al.^{[18][19]} assessed urban landscape characteristics based on the images posted on social media. Yohei Kurata et al. investigated commemorative photos shared on social media and the traits of the posters, performing a comparative analysis of tourists' photography behavior^[20]. Yoichi Ohno et al. used the data of geotagged photos posted on social media to determine the touring route within a significant Japanese garden^[21].

These explorations underscore the potential of leveraging readily available web-gathered data to reflect the possible formation of the image structure of cities, not only through firsthand experiences but also indirectly via the Internet. As online imagery proliferates rapidly, it captures the dynamic changes of image structure of cities over time. While numerous studies have utilized SNS images to outline the image structure from the perspective of the content creators^{[17][21]}, this study shifts focus towards recipients or viewers of the digital information, employing the Google Cloud Vision API (Application Programming Interface) system for content analysis.

1.4 Hypotheses and Definitions

This paper proposed the following hypotheses: 1) the contemporary recognition of cities is heavily shaped by the myriad of fragmented information and images shared by various users (senders); 2) these pieces of information crucially affect the image structure of cities, whether or not the viewers (recipients) have personally visited the city.

Concentrating on how digital image recipients perceive the city, this study defined the "image structure of cities" as the perception and imageability of cities created through the exchange of visual information about the urban environment on SNS platforms. Although this image structure exists primarily in a virtual realm, it substantially shapes how recipients experience, remember, and navigate through urban spaces in the real environment.

A glossary of other key terminologies employed in this study is listed in Table 1.

1.5 Research Flow

Firstly, this paper establishes the foundation of the research through literature review, identifies the limitations of existing research, and defines the key theoretical concepts related to the image structure of cities in the information age.

Secondly, it describes the methodologies and SNS used for

Table 1: Terminology for this research

Terminology	Definition
SNS/social media	An online platform that allows users to connect, communicate, and share information with others to build social networks and communities among individuals who share common interests, activities, backgrounds, or real-life connections
Digital information	Information shared by the general public over the Internet encompasses texts, photos, and videos uploaded by users; this study specifically focused on still visual images depicting urban environments with geographical information
Image structure	The perception and imageability of a city created by the visual information of the urban environment, as it influences how people experience, remember, and navigate through urban spaces
Viewpoint	A specific location or area within a city where a certain perspective, landscape, or scene can be observed and appreciated; these locations often serve as focal points for capturing images and can significantly contribute to the formation of urban imageability
Sender	People who upload and share digital information on SNS
Recipient	Individuals who access, read, or view the information shared on SNS, engaging with the content as consumers or audience

the survey, and illustrates the flow of data and the relationships between all proprietary software and the developed codes utilized in this paper.

Thirdly, central Tokyo's station areas are used as examples to test the methodologies for visualizing the image structure of cities. This is achieved by identifying the distribution of photo shooting locations based on geotag information from photo image data shared on social media. Along with the spatial distribution of the images, the agglomeration of viewpoints and the relationship between content and distribution patterns are analyzed.

Then, utilizing Google Cloud Vision API's image labeling function, this paper examines the content of the retrieved photo image data as a representation of the image recipients' perceptions. It also categorizes labels of image contents and identifies similarities among them. Through this analysis, it explores how the agglomerations of photos on the Internet affect social media users' perception of the city.

Finally, the research consolidates the research findings, discusses the challenges of the proposed methodologies that reveal the image structure of cities created by SNS posts, and contemplates potential prospects.

2 Research Methodologies

In this study, the data was extracted from the API provided by

Flickr, a popular photo-sharing site, to analyze the perception of the city images.

2.1 Internet-based Services Employed

2.1.1 Flickr as the Source of the City Images

Flickr, operated by Yahoo! and launched in 2004, is a popular online photo-sharing platform that allows users to upload, store, and share digital photos and videos. Flickr has now become a globally favored platform that enables users to discover photographic content through methods such as tagging and geotagging searches. The API offered by Flickr facilitates data extraction based on specified parameters and geotag information retrieval.

2.1.2 Google Cloud Platform for Image Label Analysis

In this study, we analyzed the collected photo image data, extracted the depicted urban features in each photo as text data, and utilized it for urban landscape image analysis. We employed the image label analysis service Label Annotation in Google Cloud Vision API, which capitalizes on Google's machine-learning capabilities and extensive image database to offer highly objective and replicable results. By uploading images to a dedicated server, labels conveying contents of each image can be extracted in a ranked order based on confidence scores. This process allows to analyze the aggregation of numerous images which represent image recipients' perception of the cities.

2.2 Program Development

Figure 1 illustrates the primary application integration used in this study. Additional two custom programs were also essential for data collection and image analysis.

2.2.1 Program One

We developed Program One to extract photo image data and geotag data relevant to urban landscapes, using tools such as ArcMap and ArcGIS Pro (Version 1.3 of Esri) (Fig. 2). The program set appropriate conditional parameters, such as geographical locations and time periods, to extract urban-related information available from the Flickr's API.

2.2.2 Program Two

For the examination of urban features within photos, we developed Program Two, which can transmit a pre-compiled list of images to a dedicated server to obtain analysis outcomes (Fig. 3). By utilizing a JSON file containing image URLs, we can acquire the label analysis results. For instance, a higher score from the Google Vision API indicates greater confidence in the label's accuracy. Program Two allowed for the assessment of a vast array of Flickr images, utilizing IBM SPSS Statistics 23, to determine what viewers will see on their screens. Thus, content analysis became a crucial method for understanding viewers' perceptions of images related to urban landscapes.

3 Results of Program One: Visualizing Urban Viewpoints Through Heatmaps

3.1 Study Areas

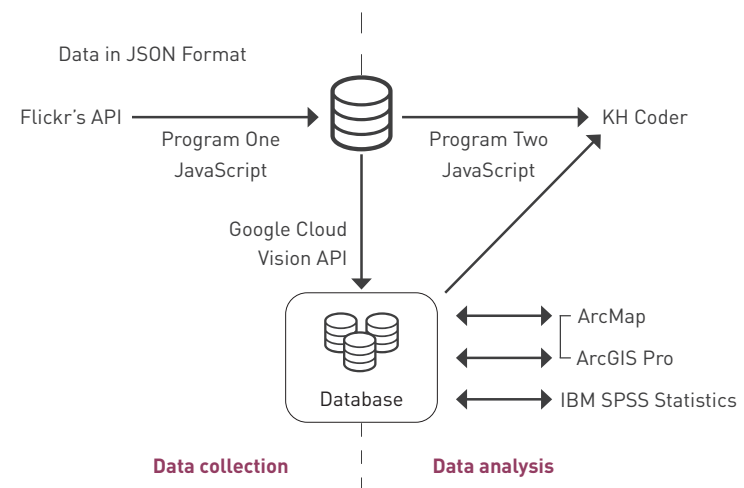
This study selected 33 major railway station areas (from both Japan Railway and Tokyo Metro) within the 23 wards of Tokyo (Fig. 4). These station areas can represent a broad spectrum of urban environments. For instance, Roppongi is recognized as an emerging commercial hub undergoing continuous redevelopment; in contrast, the Tokyo station area stands as an established business district, boasting a blend of modern high-rise structures and historic brick buildings such as the Tokyo Station Building, a registered important national treasure dating back to the mid-19th century. Several of these station areas are also notable tourist destinations. Asakusa, for example, attracts numerous tourists with its rich cultural heritage, exemplified by landmarks like Sensoji Temple and historic theaters. Additional noteworthy areas include Ginza, a high-end shopping district; Harajuku, a cultural youth culture center; and Shibuya, a major terminal station renowned

for its iconic pedestrian crossing. Station areas on the outskirts of central Tokyo, such as Nakano, Kamata, and Nippori, are lively mixed-use residential neighborhoods with beloved local shopping streets hosting family-run businesses. Beyond land use, the intricate topographical conditions further accentuate the unique characteristics of each neighborhood.

3.2 Analysis Method

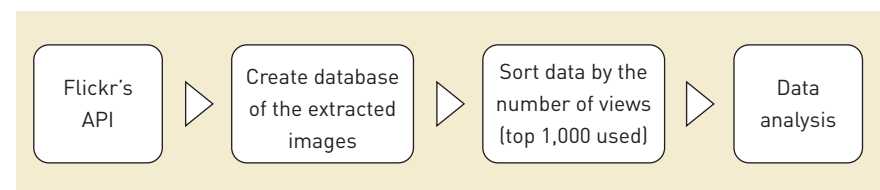
With Program One, we collected top 1,000 viewed photos with viewpoints for each area, following specific criteria that the images

1. Applications and programs used for the image analysis.
2. Data extraction flow by Program One.
3. Sample result of Label Annotation by Google Cloud Vision API.



© Tetsuya Yaguchi, Takumi Fujinuma

1

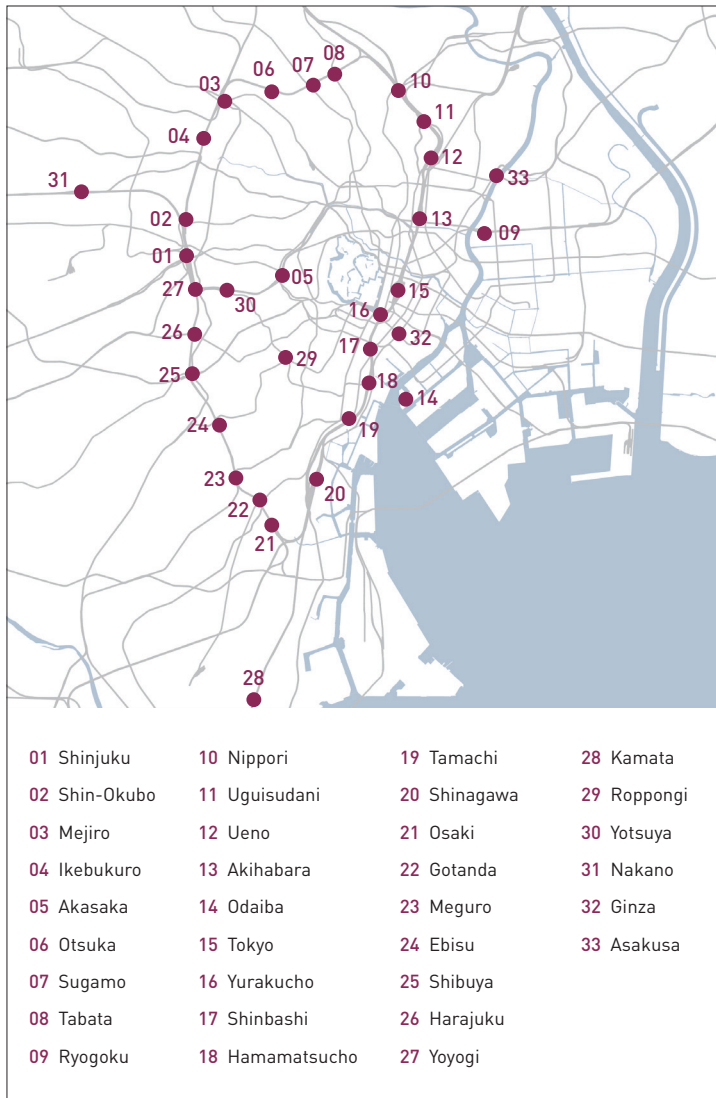


© Tetsuya Yaguchi, Takumi Fujinuma

2

Image sample from Flickr	Label	Confidence score
	Cityscape	0.83
	Light	0.72
	Night	0.72
	Tower	0.65
	Building	0.63

© Bethom33



- Distribution of the selected station areas in Tokyo.
- Density levels of viewpoints of kernel density analysis (example: Shinjuku Station Area).

must be geotagged, captured within a 1 km radius of the station, taken before September 2, 2017, and publicly accessible. As depicted in Fig. 5, we plotted the obtained geotagged data using GIS and performed a kernel density analysis to calculate the standard deviation in the density of the viewpoints for each target area. The analysis produced heatmaps that were categorized into four levels, from level 0 to level 3 by Natural Breaks/Jenks, and used the viewpoints located at levels 1 ~ 3 for detailed analysis.

3.3 Evaluation Method

The density and distribution of the viewpoints on the heatmaps were analyzed by examining the following three aspects.

3.3.1 Spatial Distribution of Viewpoints: Coverage Type/Shapes of the Viewpoint Accumulation on Heatmaps

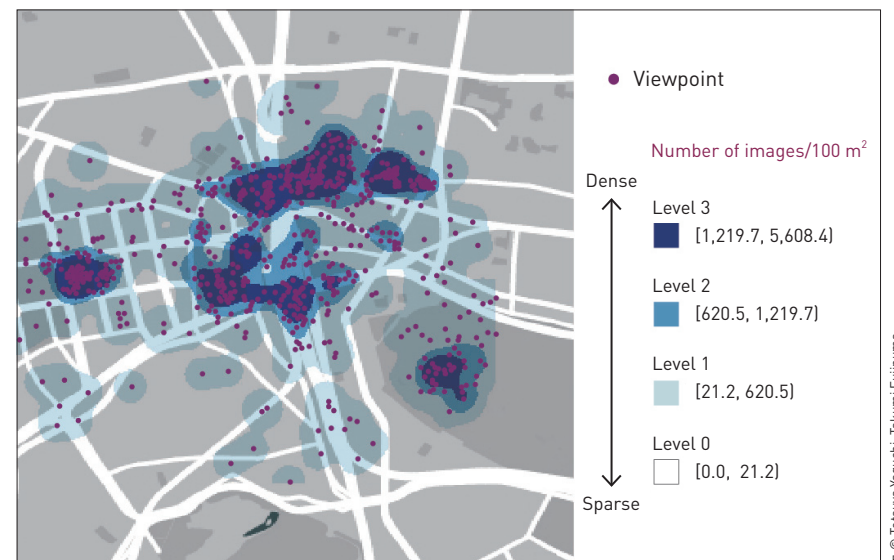
We visualized the aggregation of the viewpoints by calculating the density of geotags in photos posted on Flickr for each target area and creating corresponding heatmaps. High-density areas (level 3) indicate the presence of representative and characteristic landscapes, landmarks, or viewpoints that play a crucial role in shaping the image structure of the area.

3.3.2 Spread of Viewpoints: Coverage Rate of the Viewpoint Accumulation on Heatmaps

We assessed the dispersion of viewpoints in each target area by measuring the coverage rate of each heatmap layer within the entire map window, with level 0 representing no aggregation and level 3 signifying the densest aggregation. Utilizing the Fractal Analysis System (for Windows Ver. 3.4.7 by the National Agriculture and Food Research Organization of Japan), we calculated the coverage area for each level and their corresponding proportions. This layered representation can uncover the actual progress of urban digitalization and the distinct characteristics of each area.

3.3.3 Total Number of Images: Attractiveness Indicator

On Flickr, users can set whether to release geotags of photos and who can view them. Therefore, the count of publicly available geotagged photos on SNS that the sender wants to share with others can be used to gauge the attractiveness of the target area.



3.4 Visualization of Viewpoint Distribution

Figures 6 and 7 show the results of data extraction and kernel density analysis of all the researched station areas, respectively. Based on the configuration of the heatmaps, the target areas were classified into four coverage types.

1) Planar: This category indicated that viewpoints are broadly and continuously distributed without directionality. Major transit hubs, such as Shinjuku, Ueno, Tokyo, Shibuya, and Akihabara, fell under this category due to the dense clustering of image data. Consequently, this type tended to cover the largest area among the four categories.

2) Intersecting linear: In this category, viewpoints were distributed along intersecting lines, including the commercial streets and shopping areas adjacent to the transit stations. The vicinities of Yurakucho, Ginza, and Nippori stations exemplified this type, known for their bustling commercial streets. The linear arrangement of shopping avenues resulted in a narrower spatial

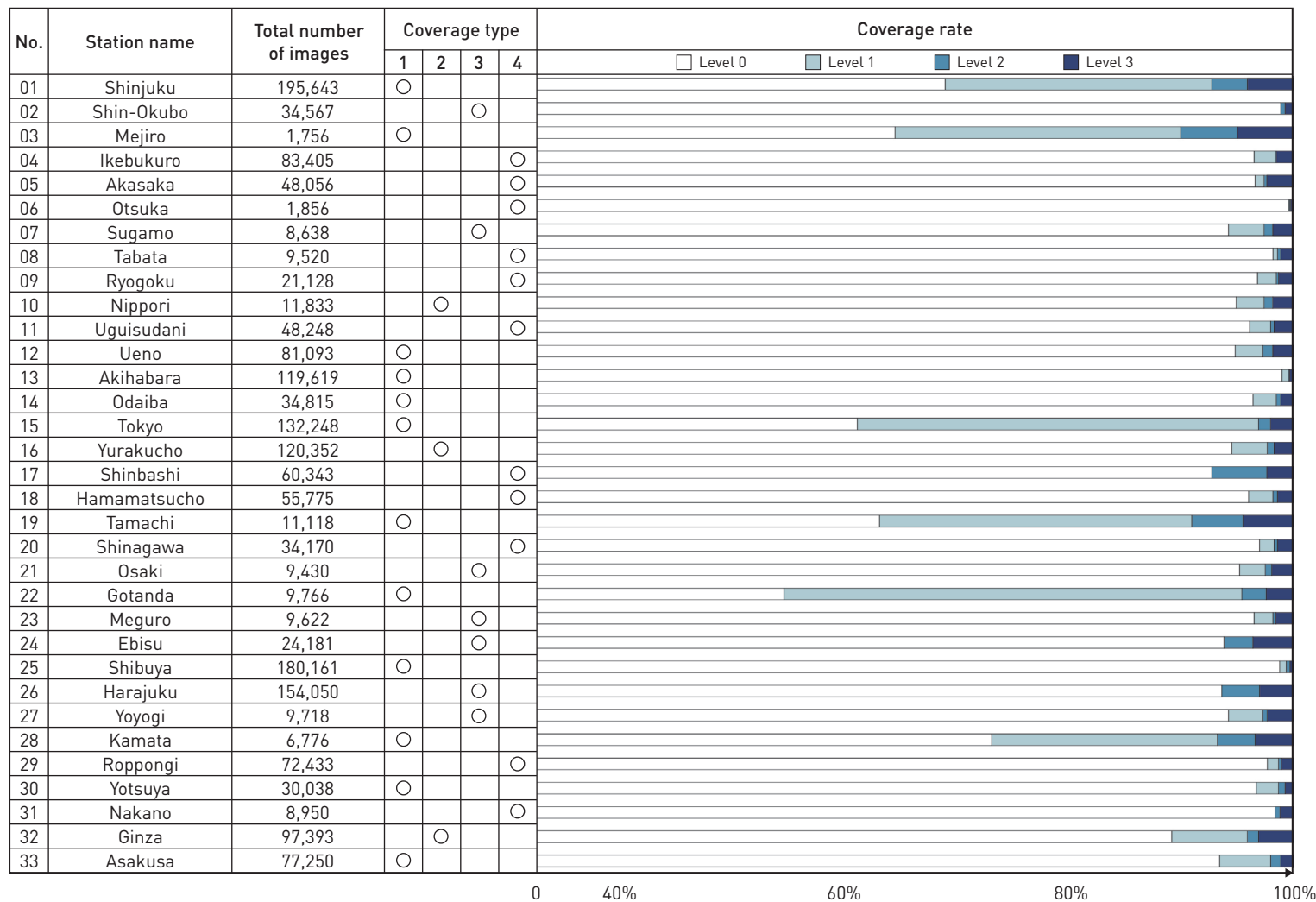
footprint for image distribution.

3) Linear: Viewpoints in this category followed a linear pattern, mirroring the linear characteristics of streets, rails, rivers, etc. Areas such as Okubo and Harajuku, which feature long, linear shopping corridors stretching from the station, and Meguro, known for its mile-long cherry blossom-lined riverfront walk, were representative of this type.

4) Nodal: This type featured viewpoints that were discrete and clustered, reflecting the view spots frequently photographed and shared, thus concentrated in compact zones. Roppongi and Ryogoku, renowned for their landmark buildings—Roppongi Hills and Ryogoku Kokugikan, respectively—fitted into this category.

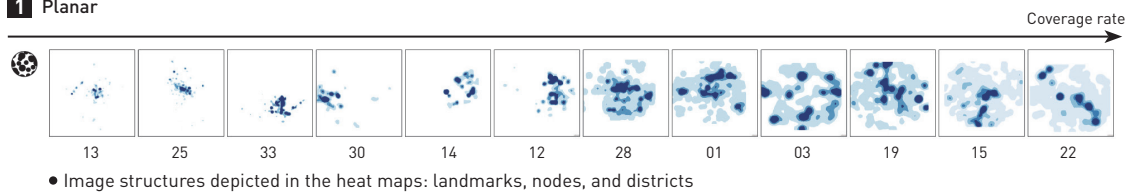
3.5 Measuring Viewpoint Coverage Rate on the Heatmaps

In this study, the coverage rate of viewpoints between level 1 (low concentration) and level 3 (high concentration) on the heatmaps were measured by the Fractal Analysis System. A



6. Data extracted from Flickr, including total number of images, coverage type, and coverage rate of the viewpoints. Coverage types 1, 2, 3, and 4 represent planar, intersecting linear, linear, and nodal, respectively.

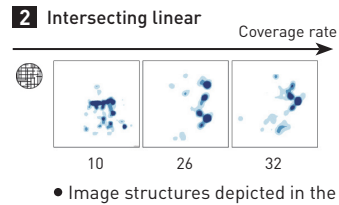
1 Planar



Sample images



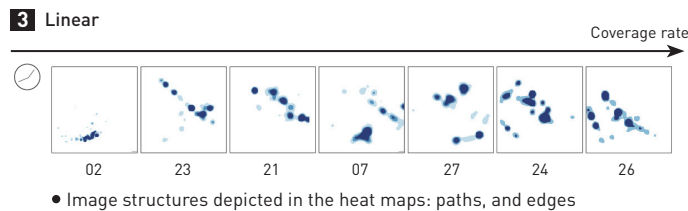
2 Intersecting linear



Sample images



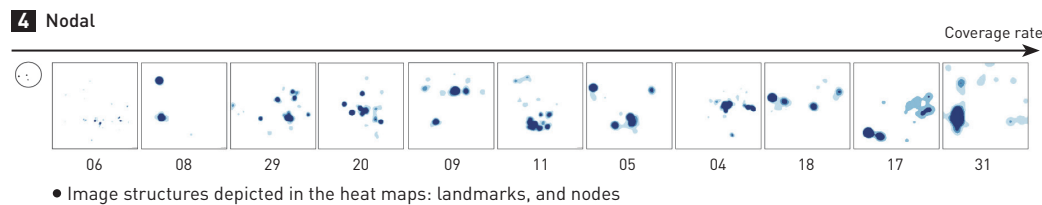
3 Linear



Sample images



4 Nodal



Sample images



7. Coverage type classification of sampled station areas.

higher proportion of level 3 indicates a more pronounced digital representation of the urban landscape (Figs. 6, 7). Areas with a level 3 coverage rate might indicate the presence of representative urban landscapes, as deduced from the clustering of posted images. When the proportion of level 3 is high, it implies that multiple viewpoints are spread across the area, exhibiting varying perspectives and distances to the same urban landscape scenes. Consequently, there might be high frequencies of the same urban landscape images in relatively larger areas, such as images of the Tokyo tower from various viewpoints. Conversely, a minimal level 3 coverage rate suggests that key viewpoints are concentrated in a specific location.

Meanwhile, areas identified as level 1, which reflect conditions with a sparse distribution of shared visual content, were found to

be usually located on the periphery of station areas or near stations with lesser foot traffic.

3.6 Total Number of Images Posted for Each Target Area

We extracted the total number of images taken during the specified period for each target area from Flickr. The data showed a range with a maximum of 195,643 and a minimum of 1,756 images.

Notably, areas frequently visited by commuters, shoppers, and tourists, including terminal stations like Tokyo, Shinjuku, and Shibuya, as well as popular tourist destinations such as Harajuku and Yurakucho, accumulated over 100,000 images within a one-kilometer radius. In contrast, stations near quieter residential neighborhoods, such as Mejiro and Otsuka, registered fewer than

2,000 images, revealing significant variations in digital content generation across different areas (Fig. 6).

The quantity of images posted by individuals indicates their inclination to share experiences^{[16][17]}. Therefore, this study posited that the concentration of images shared from the same or similar viewpoints might reflect the appeal of a location.

4 Results of Program Two: Analyzing the Contents of Extracted Photo Image Data

This step delved into the analysis of image data collected from each target area to explore which types of image content were commonly shared or exhibit notable diversity, and to discuss how this affected social media users' (image recipients') perception of the city shaped by their exposure to these shared visuals.

4.1 Analysis Overview

Utilizing Program Two, we performed a label analysis on the previous 1,000 photo samples for each target area to understand the image contents. Subsequently, KH Coder^{[22][23]} was employed to process the textual data extracted from this analysis. The Jaccard coefficients^{[22][24]} were also measured to evaluate the similarity of image contents between areas. Additionally, IBM SPSS Statistics 23 was used to conduct correlation analysis between the Jaccard coefficients and heatmap coverage rates, as well as the total number of posted images, to elucidate their relationships.

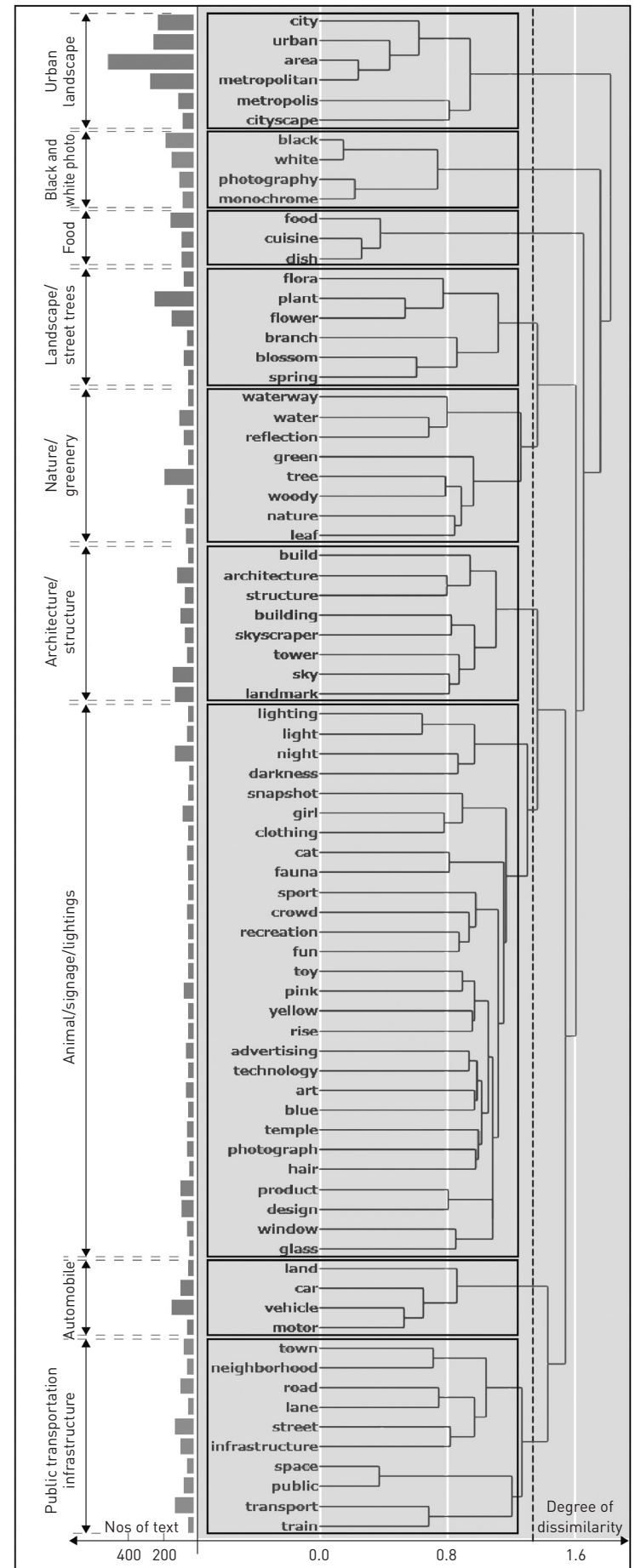
4.2 Image Label Analysis

The cluster analysis categorized these images into nine themes (Fig. 8) such as "urban landscape," "landscape/street trees," "nature/greenery," "architecture/structure," and "public transportation infrastructure." It also identified ancillary elements within the urban environment, like "food," "animal/signage/lighting," and "automobile," as labels associated with the posted images.

It is important to note that images posted on social media often carry multiple labels that provide context of the main subjects, backgrounds, and complementary objects within the images. This multiplicity reveals the complexity of urban landscape images.

4.3 Image Content Similarity Analysis

To assess the similarity of image content for each target area, we computed the Jaccard coefficients, with values ranging from 0 (no similarity) to 1 (identical). A higher value signifies a greater likeness between the image contents. Results show variances in



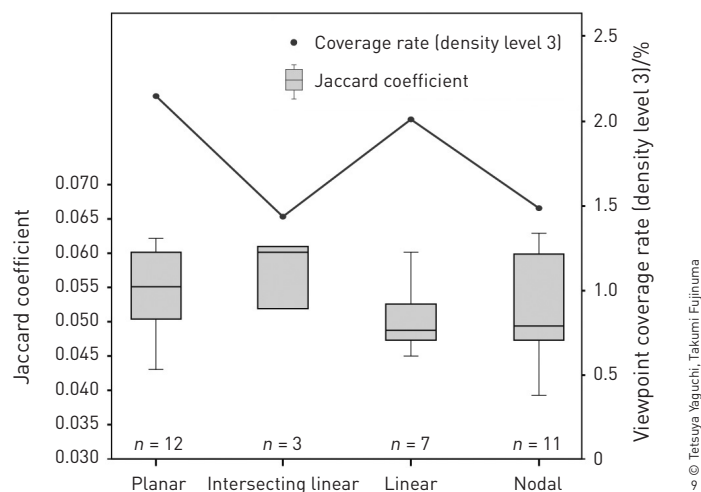
8. Result of the cluster analysis.

image content similarity across the target areas, along with notable differences tied to the distribution pattern of viewpoints (Fig. 9).

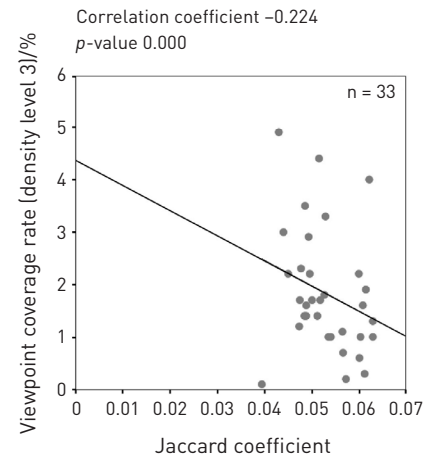
Specifically, labels derived from images of linear and nodal types, both characterized by viewpoints within a limited area and displayed a broad range of subjects (evidenced by low or median Jaccard coefficients). Areas with nodal type of the viewpoint distribution often feature unique landmarks such as museums and high-rise towers (Fig. 7), enriching the diversity of image content. In contrast, areas of intersecting linear type demonstrated a higher content similarity (indicated by the highest median Jaccard coefficient). These areas, including Yurakucho, Ginza, and Nippori, are with distinctive urban structures along single-land-use commercial streets (Fig. 7), which results in more homogeneous image labels.

4.4 Correlation Analysis

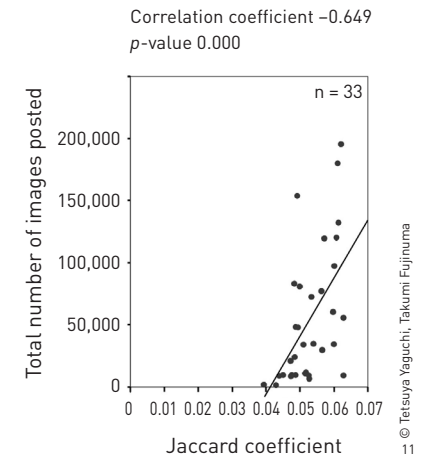
Correlation analysis was conducted using the statistical analysis software to explore the relationship between the similarity of image contents measured by Jaccard coefficients with the coverage rates at various density levels of viewpoint distribution and the total number of images (Figs. 10, 11). A weak negative correlation was observed between the Jaccard coefficient and the level 3 heatmap coverage rate (Fig. 10), while a strong positive correlation was noted between the Jaccard coefficient and the total number of images (Fig. 11). However, no correlation was found between the Jaccard coefficient and the total coverage rate (levels 1, 2, and 3 combined).



9. Image content similarity result for different viewpoint coverage types.



10. Result of correlation analysis between the Jaccard coefficient and coverage rate of the density level 3.



11. Result of correlation analysis between the Jaccard coefficient and the total number of images.

The weak negative correlation between the level 3 heatmap coverage rate and the similarity of image content can be interpreted as follows. In areas where representative viewpoints or subjects are densely scattered—such as around major terminal stations or tourist hubs, categorized in Coverage Type 1 (planar)—the contents of the shared images tend to be diverse, leading to a lower Jaccard coefficient value. In contrast, in areas where the viewpoints are narrowly confined—like shopping streets with distinctive boutiques or riverside scenes, corresponding to linear or nodal type—more uniform information tends to accumulate on social media, reflected by higher Jaccard coefficient values. In other words, in small areas with concentrated viewpoints, the digital information tends toward redundancy, which may create a stereotyped image of the city for SNS users (recipients).

It can be inferred that the presence of recognizable landmarks or themes tends to limit the variety of posts within the same area. For example, famous tourist spots like Shibuya's crossing or Shinjuku's entertainment district frequently depicted in travel guides, corresponding to planar type, are recurrently posted by numerous individuals, diminishing the overall diversity of images in that area (thus increasing the Jaccard coefficient). This phenomenon of repeated posts of similar tourist spot images, especially on social media, has been noted^[19], which may easily foster strongly stereotyped images.

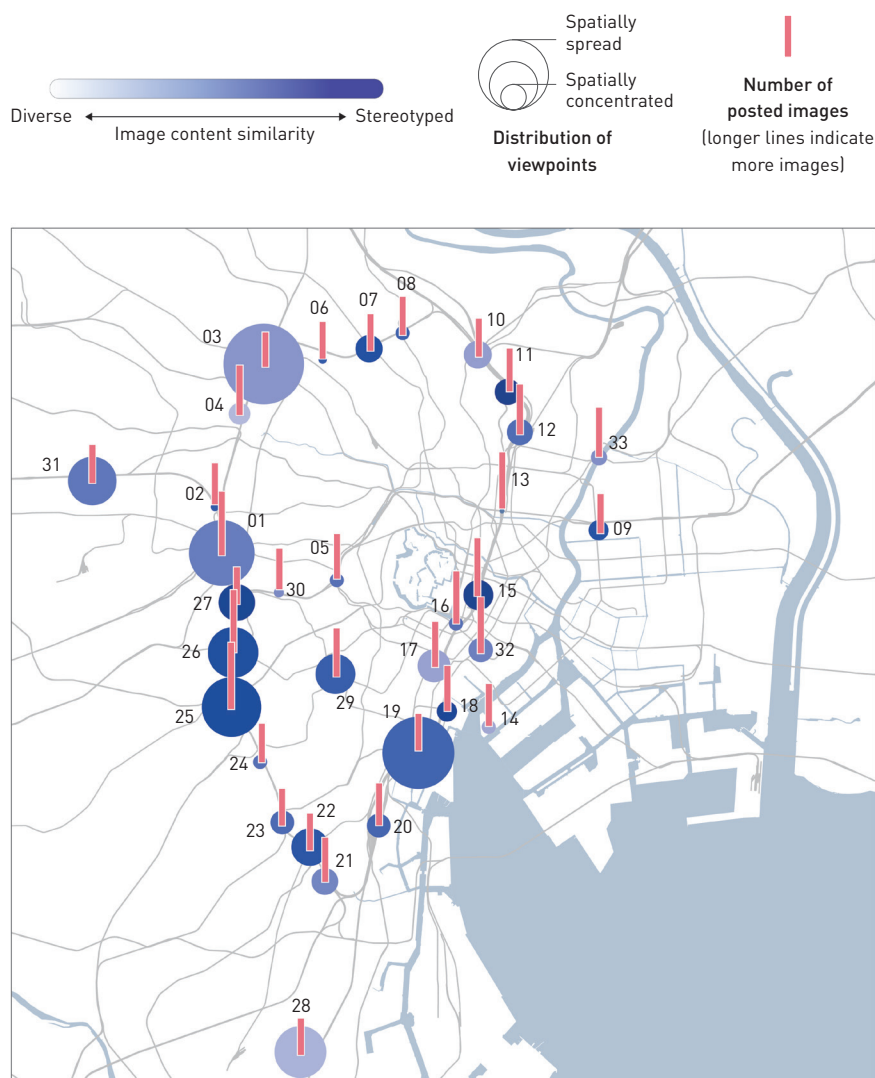
Figure 12 illustrates the relationship between the similarity/diversity of image contents, their spatial distribution of viewpoints, and the total number of photos posted on SNS.

5 Conclusions

5.1 Summary of the Study

In this paper, we analyzed the geotagged images from Flickr to map the concentration and dispersion of viewpoints throughout the urban landscapes. Through kernel density analysis, we categorized the density of viewpoints into four levels and visualized them using heatmaps, which were then further broken down into four coverage types, namely planar, intersecting linear, linear, and nodal. From the classification results and posted photo examples in Fig. 7, we inferred the urban factors influencing the viewpoint distribution, such as the presence of commercial areas, streets, rivers, and landmarks. It indicates that the distribution of images was shaped by the urban physical environment and mirrored the senders'

12. Map of relationship between image content similarity, distribution of viewpoints, and total number of images posted on Flickr.



perception of the city.

The kernel density analysis unveiled instances where viewpoints were densely clustered in relatively small areas (mainly nodal), or where a representative subject existed but was captured from multiple viewpoints, differing in perspective and distance (mainly planar). The heatmaps thus varied in depicting the concentration and dispersion of images perceived by SNS users. The study then utilized the total number of image posts to measure the attractiveness of an area.

Then, we conducted an image label analysis to represent recipients' perceptions, followed by cluster analysis and contents similarity analysis. Image labeling by Google Cloud Vision API revealed that a single image could bear multiple labels, encompassing the main subject, background, and complementary objects. The cluster analysis broadly determined how users perceive a city through the collection of images on SNS. Furthermore, the image content similarity analysis with the Jaccard coefficient indicated that the diversity of image contents was reflective of the scattered presence of unique landmarks, while the uniformity stemmed from factors such as single land usage. The correlation analysis showed that when representative viewpoints and subjects were densely and widely distributed, contents of the shared images tended to be diverse, reducing similarity. Conversely, in areas where such viewpoints and subjects were confined, social media content may become relatively homogeneous, with increased similarity. The relationship between the number of image posts on SNS and content similarity denoted that the more appealing and recognizable the landscape or subject was, the more it was posted within the same area, restricting the content diversity. On SNS, the repetitive sharing of similar images of tourist sites can lead to the formation of strongly stereotyped images.

5.2 Discussion: Potential Risk of Stereotyped Image Structure of Cities

In the era marked by advanced information and communication technologies, individuals increasingly access city-related information via computers and mobile devices, a shift from the traditional, on-site exploration. This study aimed to probe into the image structures of central Tokyo, a city that juxtaposes modern skyscrapers with traditional neighborhoods, by examining the distribution and variation of images shared on SNS. We highlighted the potential of the novel methodologies in discerning the image structure of modern cities in a state of continuous change, where both the urban environment and

public perceptions evolve. While traditional cognitive mapping methods remain valuable for understanding urban perceptions, this study underscored the significant role of digital information on social media in shaping the image structure of cities^{[25][26]}. The vast amount of readily available information on SNS sites paves the way for innovative approaches in urban studies.

However, it is also essential to understand the risk of biased image structure of cities arising on the Internet. This study observed instances where the viewpoints of the posted images were concentrated in relatively limited areas, leading to a uniform collection of images that may skew recipients' perceptions of the city from SNS. Moreover, the algorithms of existing SNS, which cater to the browsing history of users, may reinforce this bias by continuously presenting similar images, potentially triggering an echo chamber effect. It is crucial to be aware of this potential consequence^[25]. The study also noted that SNS users might conform to certain trends or behaviors fitting in or accepted by their social circles, further contributing to the proliferation of similar and stereotypical images being shared, liked, or propagated^[27].

Additionally, the study acknowledges the susceptibility of visual information to cultural and social bias^[26]. For instance, photographs posted by tourists might be influenced by their pre-acquired, edited pre-information, introducing a level of bias. Future research should strive to elucidate the differences between city image structures perceived by tourists from those of residents.

It is important to note that these factors are potential contributors to the biased image structure of the city. However, the actual impact may differ depending on many other variables, including the specific social media platform, user demographics, and individual behaviors. Attention should be paid to these differing variables when considering the broader impact and implications of this study.

5.3 Limitations and Future Research Prospects

In this study, we focused on the recipients of visual image-based digital information, specifically analyzing the geotagged images and their contents on Flickr. We leveraged the sizable amount of visual data readily available, which far exceeds the capacity of cognitive maps traditionally used since the 1960s. The scope of the exploration was limited to understanding how such visual information from SNS might facilitate recipients in constructing potential city image structures.

For future research, it is crucial to delve deeper into the relationship between image structures of cities constructed from SNS data and those formed through direct interaction with

the physical environment. Such research could examine the environmental attributes of tangible spaces with high imageability and investigate how these image structures constructed influence the behaviors of the image recipients within the city.

While Tokyo's dynamic and multifaceted nature provided a fertile ground for this study, it might be better to include more urban settings, ranging from suburban areas and rural towns to regional cities, historical towns, and cities with national heritage sites, to further validate the methodologies employed. Moreover, as SNS users continue to populate the Internet with more digital images over time, the opportunity for longitudinal analyses emerges. Incorporating a temporal dimension into the analysis will allow us to observe shifts or consistencies in the perceptions of cities over time. Gaining insights into the enduring image structures of a city can empower urban planners and designers to craft more informed visions for the future, building on the city's existing elements.

NOTE

Images from Flickr presented in this paper are licensed under Creative Commons Attribution 2.0 Generic, Attribution-ShareAlike 2.0 Generic, and Attribution-NoDerivs 2.0 Generic.

Competing interests | The authors declare that they have no competing interests.

REFERENCES

- [1] Lomborg, S., & Bechmann, A. (2014). Using APIs for data collection on social media. *The Information Society*, 30(4), 256–265.
- [2] Sevin, H. E. (2014). Understanding cities through city brands: City branding as a social and semantic network. *Cities*, (38), 47–56.
- [3] Ministry of Land, Infrastructure, Transport and Tourism of Japan. (2016). *Results of the 2015 national urban traffic characteristics survey*.
- [4] JR East Marketing & Communications, Inc. (2017). *Study finds 20-year-olds move less than 70-year-olds. The Internet accelerates the trend of “first place” and “non-movement.” Smartphone is the key to activating mobility—Move facts survey 2017*.
- [5] Flickr. (n.d.). *App garden API description documents*.
- [6] Lynch, K. (1960). *The Image of the City*. MIT Press.
- [7] Appleyard, D. (1970). Styles and methods of structuring a city. *Environment and Behavior*, 2(1), 100–117.
- [8] Kaplan, S. (1973). Cognitive Maps in Perception and Thought. In: R. M. Downs & D. Stea (Eds.), *Image and Environment: Cognitive Mapping and Spatial Behavior* (pp. 63–78). Routledge.
- [9] Carr, S., Rodwin, L., & Hack, G. (1984). Kevin Lynch—Designing the image of the city. *Journal of the American Planning Association*, 50(4), 523–525.
- [10] Nasar, J. L. (1990). The evaluative image of the city. *Journal of the American Planning Association*, 56(1), 41–53.
- [11] Hospers, G.-J. (2010). Lynch’s The Image of the City after 50 years: City marketing lessons from an urban planning classic. *European Planning Studies*, 18(12), 2073–2081.
- [12] Burgess, J. A. (1974). Stereotypes and urban images. *Area*, 6(3), 167–171.
- [13] Banai, R. (1999). A methodology for *The Image of the City*. *Environment and Planning B: Planning and Design*, 26(1), 133–144.
- [14] Filomena, G., Verstegen, J. A., & Manley, E. (2019). A computational approach to ‘The Image of the City’. *Cities*, 89, 14–25.
- [15] Liu, L., Zhou, B., Zhao, J., & Ryan, B. D. (2016). C-IMAGE: City cognitive mapping through geo-tagged photos. *GeoJournal*, 81(6), 817–861.
- [16] Okuyama, S., Inamochi, R., Nitatori, S., Shikasho, T., & Shiozaki, T. (2009). Distribution and content of photographs in Google Earth: A study on image of city in Tokyo. *Summaries of Technical Papers of Annual Meeting Architectural Institute of Japan. F-2, History and Theory of Architecture*, (2009), 757–758.
- [17] Osaki, Y., Yoshikawa, S., & Tanaka, K. (2017). Analysis and evaluation of landscape based on social media—Case studies in tourist area. *Proceedings of The City Planning Institute of Japan, Kansai Branch*, (15), 13–16.
- [18] Okutsu, H., Enta, A., Kikuchi, K., & Watanabe, H. (2010). Research on regional characteristics from photo sharing community site of web: Behavior monitoring for architectural planning on social media 1. *Summaries of Technical Papers of Annual Meeting Architectural Institute of Japan. F-1, Urban Planning, Building Economics and Housing Problems*, (2010), 1067–1068.
- [19] Okutsu, H., Enta, A., Kikuchi, K., & Watanabe, H. (2011). A study on behavior monitoring via social media 2: Extraction of hotspots based on the aggregation of actions in microblog posting locations. In: *Proceeding of the Architectural Research Meetings, Kanto Chapter, Architectural Institute of Japan*, (81), 269–272.
- [20] Kurata, Y., Mariyama, A., & Ishikawa, H. (2016). *Comparison of photography behavior in tourist spaces using Flickr images by visitor type*. In: DEIM Forum.
- [21] Ono, Y., Yoshikawa, S., & Tanaka, K. (2013). Landscape analysis in Daimyo Garden by using social media. *Papers and Proceedings of Geographic Information System Association*, (22), E-4-3.
- [22] Higuchi, K. (n.d.). *KH Coder*. KH Coder.
- [23] Manning, C. D., Raghavan, P., & Schütze, H. (2009). Stemming and Lemmatization. In: *Introduction to Information Retrieval*. Cambridge University Press.
- [24] Kamishima, T. (2003). A Survey of recent clustering methods for data mining (part 1)—Try clustering!. *Journal of the Japanese Society for Artificial Intelligence*, 18(1), 59–65.
- [25] Pariser, E. (2011). *The Filter Bubble: What the Internet Is Hiding from You*. Penguin Press.
- [26] Tian, M., Cànoves, G., Chu, Y., Font-Garolera, J., & Prat Forga, J. M. (2021). Influence of cultural background on visitor segments’ tourist destination image: A case study of Barcelona and Chinese tourists. *Land*, 10(6), 626.
- [27] Masur, P. K., DiFranzo, D., & Bazarova, N. N. (2021). Behavioral contagion on social media: Effects of social norms, design interventions, and critical media literacy on self-disclosure. *PLoS One*, 16(7), e0254670.

由社交网络服务发布内容构建的城市意象结构 ——基于Flickr的日本东京中心市区城市景观图像空间分布 及内容相似性评估

矢口哲也^{1*}, 藤沼拓巳²

*通讯作者邮箱: tetsuya.yaguchi@waseda.jp

1 日本早稻田大学创意科学与工程学院建筑系, 东京 169-8555

2 日本Goodpatch设计公司, 东京 150-0032

摘要

本研究探索了社交网络服务发布内容如何影响城市意象结构的塑造, 重点关注社交图像的空间分布及其内容相似性。有别于传统的定性研究方法, 本文借助大数据, 采用更加客观的定量方法, 勾勒出由众多社交媒体用户创建的城市意象结构。以东京中心市区为例, 本研究从Flickr平台的应用程序接口中提取了关于33个主要火车站区域的、带有地理标签图像数据, 并识别出4种视觉焦点分布类型——即平面型、交叉线型、线型和节点型, 它们分别代表了不同的城市结构。而针对图像内容的进一步分析结果表明, “城市景观”和“景观/街道树木”等主要内容显著影响了城市意象结构的形成。总体而言, 随着发布内容数量的增加和代表性视觉焦点的聚集, 用户接收到的数字信息变得更加同质化, 而这加剧了人们对城市景观的刻板印象。这些发现凸显了社交网络服务在塑造城市环境感知中的作用, 为探究由数字信息构成的城市意象结构提供了参考。

关键词

社交网络服务; 城市意象结构; Flickr; 图像分析; 刻板城市意象; 感知; 东京

文章亮点

- 利用地理信息系统 (GIS) 制图技术呈现了社交网络服务帖子所塑造的城市意象结构
- 视觉焦点的空间分布包括平面型、交叉线性型、线型和节点型4类覆盖类型
- 数字信息同质化的增加会加剧人们对城市景观的刻板印象

编辑 王颖, 马锡栋