Automated Image Generation Reflecting Current Status of PoIs for Supporting On-Site Tourist Destination Selection

Masaki Kawanaka[1], Yugo Nakamura[2], Hirohiko Suwa[1*], Keiichi Yasumoto[1*]
[1] Nara Institute of Science and Technology Ikoma, Nara, Japan
[2] Kyushu University Fukuoka, Fukuoka, Japan
[*] RIKEN Center for Advanced Intelligence Project (AIP) Tokyo, Japan

ABSTRACT

To support the decision-making process for selecting upcoming tourist spots (PoI: Points of Interest), it is necessary to share the tourist destinations' current status. Methods for sharing the current status of PoIs could include installing live cameras or having tourists share real-time photos. Still, there are challenges in terms of installation and maintenance costs and privacy concerns. In this paper, a tourism support system is proposed to aid in the decision-making of which PoI to visit next by automatically generating PoI current status images from template images and contextual information of tourist destinations while preserving privacy and presenting them to users. In the proposed system, semantic segmentation is used to divide template images into categories such as sky class, trees/grasses class, crowd class, and overall class. Based on the separately collected contextual information (weather, condition of trees/grass, crowd density, time, etc.), appropriate image transformations (using style transfer, etc) are applied to each category. The proposed system consists of an automated method for generating PoI status images and an application that presents these images to users. As part of a tourism experiment using the proposed system, 11 individuals aged 20 to 30 participated in a roughly 2-hour tour of Nara City in Japan. The results showed that compared to presenting template images, the proposed method yielded a decision-making score improvement of 0.27 and a similarity (to the current status) score improvement of 0.45 (out of five), both of which were statistically significant.

CCS CONCEPTS

• **Human-centered computing → Empirical studies in ubiquitous and mobile computing**; **Ubiquitous and mobile computing systems and tools**.

KEYWORDS

Image Generation, Current Status of Tourism Spot, Tourist Decision Support, Context-Awareness, Image Transformation

IoT 2023, November 7–10, 2023, Nagoya, Japan

ACM ISBN 979-8-4007-0854-1/23/ 11…$15.00
https://doi.org/10.1145/3627050.3627068

ACM Reference Format:
Masaki Kawanaka[1], Yugo Nakamura[2], Hirohiko Suwa[1*], Keiichi Yasumoto[1*]. 2023. Automated Image Generation Reflecting Current Status of PoIs for Supporting On-Site Tourist Destination Selection. In *The International Conference on the Internet of Things (IoT 2023), November 7–10,2023, Nagoya, Japan.* ACM, New York, NY, USA, 8 pages. https://doi.org/10.1145/3627050.3627068

1 INTRODUCTION

In recent years, systems using mobile devices such as smartphones to assist in tourism planning have been proposed [4, 6–9]. However, these tourism support systems often present users with pre-captured photos of tourist destinations, potentially leading to discrepancies in information such as seasonality and crowd levels compared to the actual site.

To address this issue, it is essential to share images reflecting the current status of tourist destinations with users. While methods like installing live cameras at these sites or having tourists share real-time photos are conceivable, they pose challenges in terms of installation/maintenance costs and privacy concerns.

Taking into account the difficulty of directly obtaining real-time images of tourist destinations, we propose a tourism support system that aids in the decision-making process of selecting the next tourist destination (hereafter called PoI: Point of Interest). The proposed system generates current status images of PoIs using pre-captured template images and contextual information, presenting them to users. Contextual information about tourist destinations, such as weather and crowd levels, constantly changes, and real-time environmental data around these sites needs to be collected. One method for real-time environmental data collection is participatory sensing [2], where tourists provide information about the current status of tourist sites. To address privacy concerns, an object recognition model [5, 10–12] running on each tourist's mobile device can be applied to images taken by tourists, extracting only the contextual information of the tourist site without sharing the images.

This paper proposes a tourism support system that automatically generates current status images of PoIs using template images and contextual information of tourist destinations while preserving privacy, presenting them to users to aid in the decision-making process of selecting the next PoI. Figure 1 provides an overview of the proposed system. The proposed system comprises an automated method for generating PoI status images and an application that presents these images to users. In the proposed method, semantic segmentation is used to divide template images into categories such as sky class, trees/grass class, crowd class, and overall class. Based on the separately collected contextual information (weather, condition of trees/grass, crowd density, time, etc.), appropriate

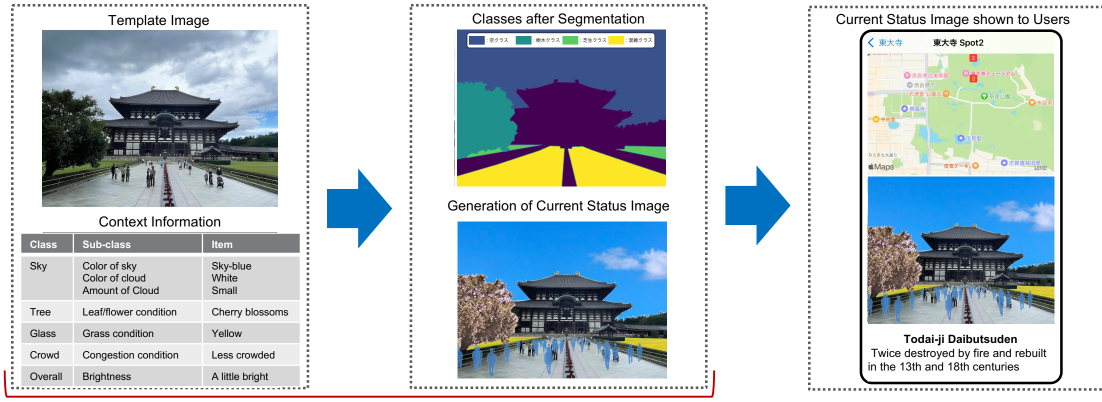


Figure 1: Overview of the proposed system

image transformations including style transfer are applied to each category.

As part of a tourism experiment using this system, 11 individuals aged 20 to 30 participated in a roughly 2-hour tour of Nara City, Japan. During the experiment, participants selected 4 PoIs to visit next out of 12 PoIs using both the conventional system (template images) and the proposed method (generated current status images). The results indicated that the proposed system, compared to the conventional system of presenting template images, improved the average decision-making score by 0.27 on a 5-point scale and the average comparison score with actual scenery by 0.45. Furthermore, the match rate between presented images and actual scenery was evaluated in terms of weather, seasonal feeling, crowd density, and time of day. As a result, compared to template images, improvements of 11.2 % in weather, 28.2 % in seasonal feeling, and 38.7 % in crowd density were observed. The match rate for time of day was 92.3 %.

2 RELATED WORK

There are numerous existing studies on tourist recommendations and travel planning support that provide information using smartphones and assist in the decision-making process of selecting the next destination spot [4, 6–9]. For instance, Hidaka et al. [4] proposed a recommendation system for onsite tourist spots considering user preferences. Additionally, CT-Planner [8] is an interactive and iterative tourism planning support system where users gradually express their preferences, and the system iteratively adjusts the plan until satisfactory.

However, in these tourism support systems, when choosing the next Point of Interest (PoI) to visit, users are presented with pre-captured template images of tourist sites. The dynamic aspects of tourism context, such as seasonality and crowd levels, known as “dynamic tourism context” [1], can differ between these template images and the actual scenery. Consequently, users might find it difficult to imagine the current status of tourist destinations, potentially making it challenging for them to select the next PoI. While methods like installing live cameras at tourist sites or having tourists share real-time photos are possible for sharing the current status of tourist destinations, they pose challenges in terms of installation/maintenance costs and privacy concerns.

Considering the difficulty of directly obtaining real-time images of tourist destinations, this study aims to develop a tourism support system that automatically generates current status images of tourist destinations and presents them to users to assist in the decision-making process of selecting the next PoI.

Table 1: Object List of Todai-ji Temple

Class	Subclass	Elements
Sky Class	Sky Color	Azure, Sky Blue, Sky Color
	Cloud Color	White Clouds, Dark Clouds
	Cloud Amount	Many, Normal, Few
Trees Class	Leaf State	Cherry Blossom, Yellow Leaves, Fresh Green, Autumn Leaves
Grasses Class	Grasses State	Fresh Green, Yellowing
Crowd Class	Congestion State	Very Crowded, Crowded, Normal, Few, Very Few
Overall Class	Brightness	Bright, Slightly Bright, Normal, Slightly Dark, Dark

3 AUTOMATIC GENERATION METHOD FOR POI CURRENT STATUS IMAGES

In our proposed method, we take pre-captured template images of PoIs and contextual information as input to generate current situation images. We utilize semantic segmentation to divide the template images into sky class, trees/grasses class, crowd class, and overall class. Then, using the segmentation results and additional contextual information acquired from the site, such as congestion, weather, and tree conditions, we perform image transformations appropriate for each class. Finally, we integrate the transformed images and adjust brightness and contrast to automatically generate current situation images of the tourist site. Hereafter, Todai-ji temple is used as an example.

3.1 List of PoI Objects

Our proposed method generates current situation images of PoIs based on contextual information. Therefore, a list of objects corresponding to the PoI's context is needed for the template images.

Figure 2: Different Views of Todai-ji Temple Captured at Different Times (Left: July 2021 and Right: April 2022)

Figure 2 shows photographs of Todai-ji temple captured at different times: on the left in July 2021 and on the right in April 2022. Differences between these photos include weather, tree and grass conditions, congestion level, and time of day. Therefore, a list of tourist site objects like the one in Table 1 can be considered, which reflects these contextual aspects. In this study, we assume that the elements associated with each subclass are represented by one-hot vectors. Hence, there is no overlap between elements like different sky colors.

We assume that the list of PoI objects can be acquired through participatory sensing using object recognition models [5, 12] operating on tourists' devices. Object recognition models have been widely used to detect objects in images, and they can accurately detect common objects such as people [10, 11].

3.2 Semantic Segmentation

In our proposed method, we apply semantic segmentation to the template images of PoIs, dividing them into multiple classes based on the PoI object list. Semantic segmentation can be accomplished using deep neural networks (DNNs) or manual methods. While DNN-based methods can automatically divide images into classes with similar features and meanings, they require a large dataset of accurately labeled images for each intended environment. Since template images are not updated frequently, we opt for manual semantic segmentation, which allows precise division into classes based on the PoI object list. Figure 3 shows the result of manually applying semantic segmentation to the Todai-ji temple template image, dividing it into sky class, trees class, grasses class, and crowd class.

Figure 3: Semantic Segmentation Results for the Todai-ji Temple Template Image

3.3 Image Transformation Based on Contextual Information

In this section, we describe the method for performing image transformations. We apply the proposed method to each class, perform image transformations using the contextual information, and then combine the transformed images into one image. Additionally, we adjust the brightness and contrast. Figure 4 presents the results of using the proposed method for image transformation. Fig. 4 (a) represents the Todai-ji temple template image. Fig. 4 (b), (c), and (d) show the results of transforming the image based on different contextual information combinations: (b) sky color + white clouds + low cloud coverage + cherry blossoms + high congestion, (c) azure color + white clouds + moderate cloud coverage + autumn leaves + moderate congestion + slightly brighter, and (d) sky color + dark clouds + high cloud coverage + yellow leaves + yellowed lawn + low congestion + slightly darker, respectively.

3.3.1 *Image Transformation for Sky Class.*

In the sky class, appropriate skies are replaced based on the conditions of clear and cloudy states obtained by the object recognition model. For this purpose, 18 sky images are prepared in advance, covering all combinations of sky color, cloud color, and cloud amount.

The results of replacing the sky in template images based on the object list of the sky class are shown in Figure 5. In Figure 5 (a), the template image of Todai-ji Temple is shown. In (b), the sky color is replaced with blue, white clouds, and low cloud density. In (c), it is replaced with azure, white clouds, and moderate cloud density. In (d), it is transformed to blue with rain clouds and high cloud density. It is evident that the impression of weather changes through the transformation of the sky class. To consider more detailed information on PoI, increasing the variety of sky images or using DNN generation models such as [14] could be considered.

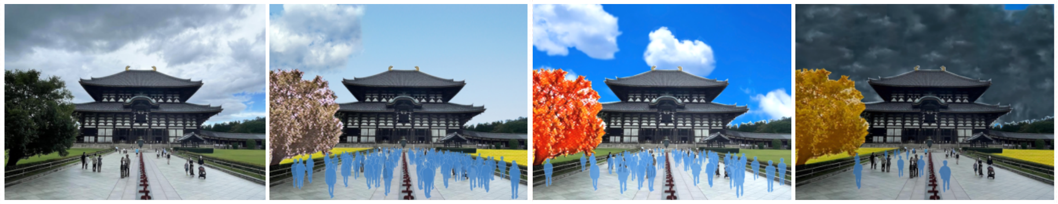
3.3.2 *Image Transformation for Trees and Grasses Classes.*

For the tree and grass classes, style transformation is performed to approximate the states of trees and grass obtained by the object recognition model. Create ML's Style Transfer is used for style transformation. Training involves 500 parameter updates, and default hyperparameters, such as Style Strength and Style Density, are used.

The results of changing the states of trees and grass in template images based on the object list of the tree and grass classes are shown in Figure 6. In Figure 6 (a), the template image is shown. In (b), it is transformed into cherry blossoms; in (c), into autumn leaves; and in (d), the grass is turned yellow. Style transformation allows maintaining the shape of trees and grass in template images while transforming them into any style.

3.3.3 *Image Transformation for Crowd Class.*

For the crowd class, image transformation is performed to simulate the levels of congestion obtained by the object recognition model. Congestion level can be represented by changes in the number of people on the space. Based on the 5 levels of congestion (very crowded, crowded, normal, few, very few) obtained by the object recognition model, silhouettes of people are overlaid onto the crowd class. Therefore, appropriate resizing of people's silhouettes needs to be done to match the template image. This is achieved by calculating the appropriate size for each coordinate based on the size of people present in the template image. The size of people present in the template image is



Template image | Sky-blue + white cloud + few clouds + cherry blossoms + grasses yellowing + very crowded | Azure color + white cloud + normal clouds + autumn leaves + normal crowds + a little bright | Sky color + black cloud + cloudy + grasses yellowing + a little dark

Figure 4: Results of Image Transformation using the Proposed Method

Template image | Sky-blue + white cloud + few clouds | Azure color + white cloud + normal clouds | Sky color + black cloud + cloudy

Figure 5: Results of Image Transformation for the Sky Class

Template image | Cherry blossoms | Autumn leaves | Grasses yellowing

Figure 6: Results of Image Transformation for Tree and Grass Classes

Template image | 150 human silhouettes | 80 human silhouettes | 10 human silhouettes

Figure 7: Results of Image Transformation for Crowd Class

Template image | bright | A little bright | A little dark

Figure 8: Results of Image Transformation for Overall Class

obtained using bounding boxes output by YOLOv3 [11]. By representing the relationship between the size of bounding boxes and the y-coordinate of people at different positions, a linear function expresses the relationship between the y-coordinate and the size of people.

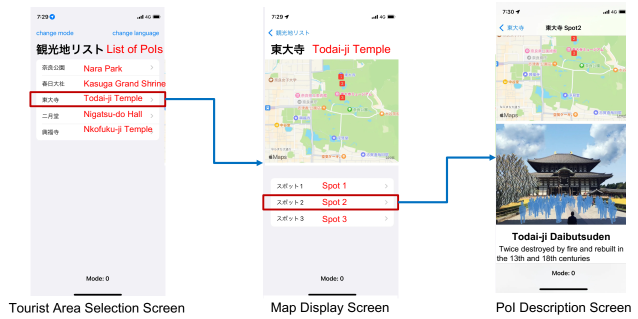


Figure 9: Application Screens

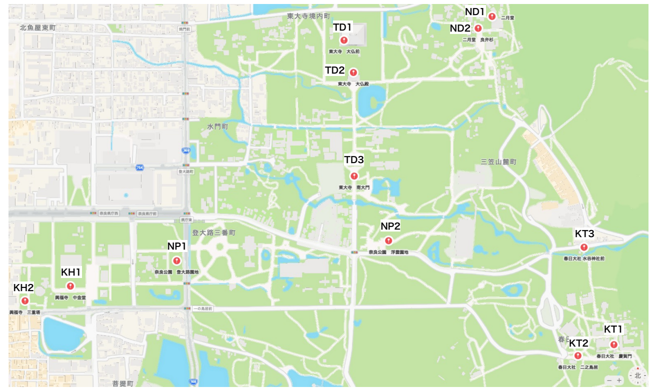


Figure 10: Target PoI locations

The results of changing the congestion state of template images based on the object list of the crowd class and the relationship between bounding box size and y-coordinate are shown in Figure 7. In Figure 7(a), the template image is shown. In (b), 150 people are added, in (c), 80 people are added, and in (d), 10 people are added. The results show that appropriate overlaying with the right size can be achieved based on the placement of people.

3.3.4 Image Transformation for Overall Class. For the overall class, image transformation is performed to reflect the current time of day based on the brightness information obtained by the object recognition model. Using OpenCV's convertScaleAbs, brightness and contrast are adjusted.

The results of changing the brightness of template images based on the object list of the overall class are shown in Figure 8. In Figure 8 (a), the template image is shown. In (b), it is made brighter, in (c), slightly brighter, and in (d), slightly darker. Darkening the brightness allows for an impression of evening or night, simulating a sunset. However, to distinguish between morning and midday, additional elements beyond just brightness adjustments might be needed.

4 MOBILE APPLICATION

In this section, we implement a mobile application that presents the current status images generated using the proposed method. The application is developed as an iOS app, and Swift is used as the programming language. The purpose of this application is to aid in the selection of the next Point of Interest (PoI) by providing users with current status images generated using the proposed method, in addition to PoI location information and descriptions.

For this reason, the implemented mobile application should meet two requirements: (i) the ability to select a tourist area and display its PoIs on a map, and (ii) the ability to present the current status image of a PoI when selected. The user interface of the mobile application is depicted in Figure 9.

The subsequent sections will provide detailed information about each screen of the application.

4.1 Sightseeing area selection screen

The screen in the left section of Figure 9 displays the tourist area options. By selecting a tourist area, users transition to the map display of PoIs, shown in the center of Figure 9. The tourist areas targeted by this application are "Nara Park," "Kasuga Grand Shrine," "Todai-ji Temple," "Nigatsu-do Hall," and "Kofuku-ji Temple."

4.2 PoI map displaying screen

The screen depicted in the center of Figure 9 displays the map with the PoIs of the selected tourist area. On the map, each PoI is represented by a map pin with a number corresponding to its identifier. Additionally, the user's current location using the application is displayed on the map. The spatial arrangement of the PoIs targeted by this application is shown in Figure 10, and their names are listed in Table 2.

The application focuses on tourist areas within Nara City, Nara Prefecture in Japan, and provides the opportunity to select from among 12 PoIs. To access more details about a particular PoI, users can select the PoI button located beneath the map, leading to the detailed PoI screen on the right side of Figure 9.

4.3 Detailed screen of each PoIP

The screen shown on the right side of Figure 9 is the detail view of a selected PoI. This screen presents the PoI's location on the map along with the generated current image of the PoI using the proposed method. Additionally, a brief description of the PoI is provided. It is important to note that the context reflected in the current image varies for each PoI, as indicated in Table 2.

5 EXPERIMENTS

This section evaluates the practicality and effectiveness of the proposed system and presents the results of the evaluation experiments.

5.1 Experiment Overview

The goal of this experiment is to investigate whether the proposed method, which involves presenting users with generated current images using the proposed technique, can support decision-making regarding the next tourist spot to visit. To achieve this, a comparison is made between presenting current images generated using the proposed method and presenting template images of tourist spots to the users. The experiment aims to determine whether the

Table 2: Tourist areas, PoIs and classes to be reflected

Tourist Area	Tourist Spot Name	Symbol	Sky class	Tree class	Grass class	Crowd class	Overall class
Nara Park	Todai-ji Nigatsu-do Park	NP1	✓	-	✓	✓	✓
	Todai-ji Ukimido Park	NP2	✓	-	✓	✓	✓
Kasuga Grand Shrine	Kasuga Grand Shrine Keiga Gate	KT1	✓	✓	-	✓	✓
	Kasuga Grand Shrine Ni-no-torii Gate	KT2	✓	-	-	✓	✓
	Kasuga Grand Shrine Mizutani Shrine	KT3	-	-	-	✓	✓
Todai-ji Temple	Todai-ji Front of Daibutsuden	TD1	-	-	-	✓	-
	Todai-ji Daibutsuden	TD2	✓	✓	✓	✓	✓
	Todai-ji Nandaimon Gate	TD3	✓	-	-	✓	✓
Nigatsu-do Hall	Nigatsu-do Hall	ND1	✓	-	-	✓	✓
	Nigatsu-do Hall Ryoben Cedar	ND2	✓	-	-	✓	✓
Kofuku-ji Temple	Kofuku-ji Chukondo Hall	KH1	✓	-	✓	✓	✓
	Kofuku-ji Sanju-no-to	KH2	✓	-	✓	✓	✓

proposed method enhances the decision-making process. Additionally, after the experiment, a questionnaire is conducted to assess the practicality of the proposed system.

This experiment was conducted with a target group of 11 participants (8 males and 3 females) in their 20s to 30s, with an average age of 24.5 and a standard deviation of 3.39. The experiment took place at 12 Points of Interest (PoI) located in Nara City. Participants were provided with an explanation of the experiment in advance and installed the proposed application on iOS devices.

The experiment compared two methods: one displaying real-time context-generated images using the proposed approach (Proposed System), and the other showing template images of tourist spots (Conventional System). The aim was to assess whether these methods could support the decision-making process for the next PoI to visit and to determine whether the provided images matched the actual scenery.

The experiment was conducted on Saturday, January 21, 2023, starting at 14:00. Participants walked to 8 PoIs during the experiment. The weather was cloudy, with a temperature of approximately 5.9°C, and no special events were taking place in the vicinity.

Initially, the 11 participants selected 4 PoIs for visiting by viewing the template images of tourist spots using the conventional system. Subsequently, they selected 4 PoIs for visiting by viewing the context-generated images produced using the proposed system. To generate the current status images, context information was essential. The experiment's assistants provided photos of each PoI's surroundings every 15 minutes. These photos were used as a basis for updating the current status images of the tourist spots.

During the course of the tourist activities, participants were asked to respond to a questionnaire that inquired whether the provided images were helpful for decision-making and how they felt about the correspondence between the presented images and the actual scenery at each visited spot. After concluding the tourist activities, a follow-up questionnaire was administered to confirm the practicality of the system.

5.2 Questionnaire Content

In this experiment, two questionnaires were employed: a questionnaire during the tourist activities and a questionnaire after the tourist activities. During the tourist activities, participants were asked to compare the effectiveness for decision-making (Decision-Making Score) and the similarity between the provided images and the actual scenery of each PoI (Similarity Score) after the completion of visiting each PoI. The questionnaire prompts are presented in Table 3.

The comparison between the images provided by the application and the actual scenery was conducted by juxtaposing the images displayed when selecting the next PoI to visit and the real-life surroundings. To explore the distinctions in the comparison between the images provided by the application and the actual scenery, these images were collected during the aforementioned questionnaire sessions. Additionally, for the closed-ended questions, participants were asked to provide free-text descriptions of any discrepancies if they answered "No" to the questions.

Upon concluding the tourist activities, a System Usability Scale (SUS) questionnaire with ten questions was administered to evaluate the usability of the application as part of assessing its practicality.

5.3 Results of the Sightseeing Experiment

The number of visitors to each PoI during this experiment is shown in Table 4. From Table 4, it can be observed that both the conventional system and the proposed system had at least one visitor for each PoI. Regarding the number of visitors, except for the Nigatsu-do Hall, there is little difference between the conventional system and the proposed system. The reason for the lower number of visitors to Nigatsu-do Hall in the conventional system is likely due to its relatively distant location from the starting point of the experiment at Kintetsu Nara Station.

Next, the results of the impact of PoI selection decision based on the presented images and the comparison with the actual scenery are shown in Table 5. First, the average decision-making score on a 5-point scale was 3.98 for the conventional system and 4.25 for the proposed system, both of which indicate high scores. This suggests that the decision-making process for selecting the next PoI benefits from both linguistic information and visual imagery. Furthermore, the proposed system showed a higher average decision-making score by 0.27 compared to the conventional system. This indicates that presenting context-aware current images generated using both context information and PoI template images is more helpful for the decision-making process than providing only template images.

Table 3: Questionnaire Content

Questionnaire Type	Question
5-point Likert Scale	Q1: Was the presented image helpful for decision-making? Q2: How do you compare the presented image with the actual scene?
Closed Questionnaire	Q3: Did the presented image match the actual weather? Q4: Did the presented image match the actual sense of season? Q5: Did the presented image match the actual crowd level? Q6: Did the presented image match the actual time of day?

Table 4: System-wise Visitors to Each PoI

Spot Name	Nara Park		Kasuga Shurine			Todaiji Temple			Nigatsudo Hall		Kofukuji Temple	
Spot Name	NP1	NP2	KT1	KT2	KT3	TD1	TD2	TD3	ND1	ND2	KH1	KH2
Visitors using conventional system	5	5	5	5	5	1	2	4	1	2	5	4
Visitors using the proposed system	5	5	3	3	4	1	1	6	4	5	4	3

Table 5: Comparison Results between the Influence of PoI Selection Decision in Presented Images and Actual Scenes

	Conventional System	Proposed System	p-value	Significance ($p \leq 0.05$)
Mean Decision Score	3.98	**4.25**	**6.70×10^{-4}**	✓
Standard Deviation of Decision Score	0.72	0.83		
Mean Similarity Score	3.66	**4.11**	**3.87×10^{-6}**	✓
Standard Deviation of Similarity Score	1.15	1.03		
Proportion Answered with Matching Weather	0.63	**0.74**	1.44×10^{-1}	-
Proportion Answered with Matching Seasonality	0.39	**0.67**	**2.72×10^{-2}**	✓
Proportion Answered with Matching Congestion	0.48	**0.86**	**5.78×10^{-5}**	✓
Proportion Answered with Matching Time of Day	0.92	0.92	5.00×10^{-1}	-

Next, the average similarity score, which compares the provided images with the actual scenery, was 3.66 for the conventional system and 4.11 for the proposed system, with the latter being higher by 0.45. This implies that the proposed method effectively conveys the current states of the sightseeing spots.

Additionally, the normality of the decision-making scores and the similarity scores between the proposed and conventional systems was tested using the Shapiro-Wilk test, revealing that none of these scores followed a normal distribution. Consequently, the non-parametric Wilcoxon signed-rank test was employed to compare the decision-making scores and the similarity scores.

The results of this test showed a significant difference, with a p-value of 6.70×10^{-4} for the decision-making scores, and a p-value of 3.87×10^{-6} for the similarity scores, confirming that the proposed system significantly outperformed the conventional system in terms of supporting the decision-making process and providing imagery that accurately reflects the current state of the sightseeing spots.

Finally, the degree of matching between the actual scenery and the provided images in terms of various context factors was compared. When comparing the weather in the presented images with the actual weather, the percentage of respondents indicating a match increased by 11 % for the proposed system compared to the conventional system.

Furthermore, the reason for the 63 % match rate when presenting template images is that some users indicated a match even when the actual weather conditions and those under which the template images were captured were different. This could be due to the presence of users who perceived a match in cases where the amount

Figure 11: Photos of the template image and the actual scenery when weather-matched responses were given

of clouds differed, as shown in Figure 11. However, some users also responded that the weather was different based on cloud cover, suggesting the necessity for image transformation for different sky conditions to accurately convey the current situation to all users.

In addition, when comparing the presented images with the actual scenery in terms of seasonal feel, the percentage of responses indicating a match increased by 28 % for the proposed system compared to the conventional system. Users who responded “No” to the seasonal feel in the proposed system pointed out differences in color and quantity of leaves and grass. To achieve higher accuracy in reflecting seasonal feel, it is apparent that improving the quality of the current images generated using the proposed method, such as utilizing other style transfer techniques, is required.

Moreover, when comparing the presented images with the actual scenery in terms of congestion level, the proportion of respondents indicating a match was 86 % for the proposed system, showing a 38 % increase compared to the conventional system. The proposed system effectively represented congestion levels by overlaying silhouettes of people.

Lastly, when comparing the time of day in the presented images with the actual time, the match rate was 92 % for both the proposed and conventional systems. This similarity in time distribution was due to many template images being captured during a similar timeframe as the current experiment.

Using a one-sided Z-test, the matching rates for various context factors in the presented images from the proposed and conventional systems were tested. The results of these tests showed significant differences in the matching rates for the seasonal feel and congestion level, with p-values of 2.72×10^{-2} and 5.78×10^{-5}, respectively. On the other hand, the p-values for weather and time of day were 1.44×10^{-1} and 5.00×10^{-1}, respectively, indicating no significant differences. The current representation of weather involves replacing the sky in the template images with pre-prepared sky images. Thus, if suitable sky images that resemble the actual sky cannot be prepared in advance, it could lead to results where no significant difference is observed. To address this issue, using generative models like GANs [3] to generate sky images might be necessary.

After the experiment, the usability of the application was assessed through a 10-question System Usability Scale (SUS) questionnaire answered by the experimental subjects. The results yielded a SUS score of **84.1**, which is relatively high.

6 CONLUSIONS

In this paper, we proposed a tourism support system that assists in making decisions about the next Point of Interest (PoI) to visit by presenting current status images generated using an automated method. In the experimental validation of the proposed tourism support system, the average decision score was 4.25, indicating that image information is crucial for PoI selection decisions. Furthermore, the average comparison score between the presented images and the actual scenery was 4.11, demonstrating that the proposed system adequately represented the current state of the tourist locations.

When comparing the proposed system to the conventional system, the average decision score increased by 0.27, and the average comparison score with the actual scenery improved by 0.45. Additionally, evaluating whether the presented images matched the actual scenery in terms of weather, seasonal feel, congestion level, and time of day revealed that compared to the template images, the current images achieved an 11 % improvement in matching for weather, 28 % for seasonal feel, and 38 % for congestion level. Moreover, regarding the time of day, both the template images and the current images achieved a high matching rate of 92 %.

It is important to note that the proposed system generates the current images of tourist locations under the assumption that the latest contextual information is accessible. For future practical applications, it will be necessary to investigate the feasibility of obtaining the latest contextual information about tourist locations by applying photos taken by tourists to object recognition models operating on mobile devices[5, 10–13].

Acknowledgment

This work was supported by JSPS Kakenhi Grant Number JP21H03431.

REFERENCES

[1] Gregory D. Abowd, Anind K. Dey, Peter J. Brown, Nigel Davies, Mark Smith, and Pete Steggles. 1999. Towards a Better Understanding of Context and Context-Awareness. In *Handheld and Ubiquitous Computing*, Hans-W. Gellersen (Ed.). Berlin, Heidelberg, 304–307.

[2] Jeffrey A Burke, Deborah Estrin, Mark Hansen, Andrew Parker, Nithya Ramanathan, Sasank Reddy, and Mani B Srivastava. 2006. Participatory sensing. *Center for Embedded Network Sensing* (2006).

[3] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. 2014. Generative adversarial nets. *Advances in neural information processing systems* 27 (2014).

[4] Masato Hidaka, Yuki Kanaya, Shogo Kawanaka, Yuki Matsuda, Yugo Nakamura, Hirohiko Suwa, Manato Fujimoto, Yutaka Arakawa, and Keiichi Yasumoto. 2020. On-site Trip Planning Support System Based on Dynamic Information on Tourism Spots. *Smart Cities* 3, 2 (2020), 212–231.

[5] Andrew G. Howard, Menglong Zhu, Bo Chen, Dmitry Kalenichenko, Weijun Wang, Tobias Weyand, Marco Andreetto, and Hartwig Adam. 2017. MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications. https://doi.org/10.48550/ARXIV.1704.04861

[6] Shogo Isoda, Masato Hidaka, Yuki Matsuda, Hirohiko Suwa, and Keiichi Yasumoto. 2020. Timeliness-Aware On-Site Planning Method for Tour Navigation. *Smart Cities* 3, 4 (2020), 1383–1404.

[7] Shogo Kawanaka, Juliana Miehle, Yuki Matsuda, Hirohiko Suwa, Wolfgang Minker, and Keiichi Yasumoto. 2021. Task allocation interface design and personalization in gamified participatory sensing for tourism. *arXiv* 2105.12032 (2021), 1–21. https://arxiv.org/abs/2105.12032

[8] Yohei Kurata, Yasutaka Shinagawa, and Tatsunori Hara. 2015. CT-Planner5: a Computer-Aided Tour Planning Service Which Profits Both Tourists and Destinations.

[9] Shogo Matsuno, Reiji Suzumura, and Minoru Ohyama. 2018. Tourist Support System Using User Context Obtained from a Personal Information Device. In *Adjunct Publication of the 26th Conference on User Modeling, Adaptation and Personalization* (Singapore, Singapore) *(UMAP '18)*. 91–95. https://doi.org/10.1145/3213586.3226217

[10] Joseph Redmon, Santosh Divvala, Ross Girshick, and Ali Farhadi. 2016. You only look once: Unified, real-time object detection. In *Proceedings of the IEEE conference on computer vision and pattern recognition*. 779–788.

[11] Joseph Redmon and Ali Farhadi. 2018. Yolov3: An incremental improvement. *arXiv preprint arXiv:1804.02767* (2018).

[12] Mark Sandler, Andrew Howard, Menglong Zhu, Andrey Zhmoginov, and Liang-Chieh Chen. 2018. MobileNetV2: Inverted Residuals and Linear Bottlenecks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*.

[13] Shusaku Tomita, Jose Paolo Talusan, Yugo Nakamura, Hirohiko Suwa, and Keiichi Yasumoto. 2022. FedTour: Participatory Federated Learning of Tourism Object Recognition Models with Minimal Parameter Exchanges between User Devices. In *2022 IEEE International Conference on Pervasive Computing and Communications Workshops and other Affiliated Events (PerCom Workshops)*. 667–673. https://doi.org/10.1109/PerComWorkshops53856.2022.9767391

[14] Zhengxia Zou. 2020. Castle in the Sky: Dynamic Sky Replacement and Harmonization in Videos. *arXiv preprint arXiv:2010.11800*.