Finding “Retro” Places in Japan: Crowd-sourced Urban Ambience Estimation

Shu Anzai
anzai.shu.an4@is.naist.jp
Nara Institute of Science and
Technology
Ikoma, Nara, Japan

Taichi Murayama
murayama.taichi.mk1@is.naist.jp
Nara Institute of Science and
Technology
Ikoma, Nara, Japan

Shuntaro Yada
s-yada@is.naist.jp
Nara Institute of Science and
Technology
Ikoma, Nara, Japan

Shoko Wakayama
wakamiya@is.naist.jp
Nara Institute of Science and
Technology
Ikoma, Nara, Japan

Eiji Aramaki
aramaki@is.naist.jp
Nara Institute of Science and
Technology
Ikoma, Nara, Japan

ABSTRACT

Understanding the ambience of an area is essential for making various geographical decisions. This kind of ambience (e.g., “beautiful,” “quiet,” “happy,” “retro,” etc.) is due to not only physical features, such as scenery and functionality, but also subjective information based on people’s perceptions and experiences. Although scholars in various fields have attempted to quantitatively measure subjective human experiences, the amount of location-based subjective data varies depending on the area. Thus, we have proposed a method to quantify individuals’ perceptions of urban ambiances, as well as a means to anticipate these perceptions from landscape images. In this study, we have focused on the “retro” ambience of old Japanese cities. First, we used crowdsourcing to assign ambience scores to the cities of Nara and Kyoto, the ancient capitals of Japan, based on the level of perception of retro people in the landscape images. Further, we trained a deep learning model to estimate the ambience score of unknown landscape images using previously labelled data. Finally, we discussed the utility of our method for routing applications and exploring its generalizability across different cities.

CCS CONCEPTS

• **Information systems** → **Location based services**; • **Human-centered computing** → **Empirical studies in collaborative and social computing**.

KEYWORDS

Urban Informatics, Crowdsourcing, Social Sensor, Deep Learning, Tourism Information, Route Recommendation

ACM Reference Format:

Shu Anzai, Taichi Murayama, Shuntaro Yada, Shoko Wakayama, and Eiji Aramaki. 2021. Finding “Retro” Places in Japan: Crowd-sourced Urban Ambience Estimation. In *5th ACM SIGSPATIAL International Workshop on Location-Based Recommendations, Geosocial Networks, and Geoadvertising (LocalRec’21)*, November 2–5, 2021, Beijing, China. ACM, New York, NY, USA, 4 pages. <https://doi.org/10.1145/3486183.3490999>

1 INTRODUCTION

Geographic information is important in the field of tourism and transportation as it is often used for route recommendations and city recognition. This includes the geographic properties of the ambience, or the affective quality of a place (e.g., “beautiful,” “quiet,” “happy,” “retro,” etc.). Ambience does encompass not only the physical, or functional features of a landscape, but it also derives from the subjective perceptions and experiences of the people with regard to that landscape. Despite this fact, previous studies have focused primarily on the physical components of the ambience, hence overlooking the subjective components.

For instance, Wakamiya et al. [2] proposed a path finding method based on the color brightness of Google Street View landscape images. However, a study on the Happy Map project [1] is an exception, as it is closely related to our study. Here, the authors quantified the ambience of London in terms of three indices—“beautiful,” “quiet,” and “happy”—and proposed a route based on the ambience. However, this method cannot be applied to areas where the ambience has not been labeled through crowdsourcing or Flickr tags. Moreover, it is difficult to comprehensively collect people’s subjective perceptions of urban ambience.

In this study, we aim to fill this gap by (1) crowdsourcing labels to quantify the ambience of a city based on people’s impressions of street and landscape images, and (2) investigating a method to estimate the affective ambience from such landscape images. In our case study, we focused on quantifying “retro” ambiances using landscape images of Nara and Kyoto, two of the most retro-rich cities in Japan. We chose this attribute as retro is considerably significant in the field of tourism. Previous studies [3] have demonstrated that historic landscapes can stimulate tourism demand. For example, presenting retro information to visitors in a tourist attraction may modify the visitor’s behavior. The degree to which people

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

LocalRec’21, November 2–5, 2021, Beijing, China

© 2021 Association for Computing Machinery.

ACM ISBN 978-1-4503-9100-9/21/11...\$15.00

<https://doi.org/10.1145/3486183.3490999>

Figure 1: A part of the crowdsourcing-based map in the city of Nara. Red points in the map show intersections. The darker the color, the greater the perceived retro of the street.

perceive a landscape as retro is an important indicator. Each landscape image is collected from Google Street View, as its service covers a wide range of streets in the target cities. We adapted Dorsch et al.'s method [6] for recognizing cities using deep learning to train a deep learning model to estimate the ambience of unlabeled landscape images. Finally, we discussed the results of our experiment and outlined relevant future direction, including ways to improve the accuracy of the model.

2 PROPOSED METHOD

2.1 Dataset

As a case study, we focused on the landscapes of Nara and Kyoto. We first selected 85 streets between intersections in the city of Nara¹. For Kyoto, we divided one 2 km × 3 km area into 100 cells of 150 m × 400 m units. Then, each street or cell was divided into three equally distanced points. For each of these three points, we collected four outward-facing images from different vantages using Google Street View Static API¹. As a result, we obtained 12 images per street, or cell. In total, the dataset consists of 2,220 images: 1,020 from Nara and 1,200 from Kyoto.

To collect subjective impressions of each street, we recruited crowd workers from Yahoo! JAPAN crowdsourcing² who had previously visited each city. These crowd workers were shown a map that highlighted a specific street or cell, while providing Google Street View links that jump to the scenery at each highlighted location, and further asked to rate the retro of the streets on a four-point scale. We obtained 1,133 responses from this process, excluding inappropriate responses (e.g., the same answer is given for all questions) from crowdsourcing. Thereafter, we applied z-normalization to calculate the average score of each street, yielding the ambience score. Figure 1 illustrates some of the crowdsourcing results as a map. The darker the color, the greater the perceived retro of the street.

2.2 Model

The main goal of our model is to predict the ambience score of a street by extracting and evaluating the features of 12 landscape images from that street. To build the model, we combined two

Table 1: Estimation performance of models

Model	Correlation	MAE
Nara	0.521	0.586
Kyoto	0.342	0.718

machine learning models, a pre-trained convolutional neural network (CNN) model, and a linear regression model. For the CNN model, we used GoogLeNet [5], a 22-layer CNN that was trained on 1.8 million landscape images from the Places365³ dataset. Including this pre-trained model as a component in our own model allowed us to compensate for our limited resources. Consequently, the GoogLeNet component of our model outputs 12,000 (1,000 dimensional features × 12 images) dimensions as features. The linear regression model component obtained 12,000-dimensions from GoogLeNet as inputs, and outputs the estimated ambience score. We used the above procedure to build a model to estimate the ambience conditions from the landscape images.

3 EXPERIMENTS

We evaluated the proposed method using two experiments: ambience estimation and routing. The former directly measures the ability of the model to successfully estimate the level of retro in a location, whereas the latter demonstrates how our model would function in routing in Nara.

3.1 Ambience Estimation

For the estimation, we split the dataset into a 70:30 (training data: test data) ratio. For evaluation metrics, we used correlation and mean absolute error (MAE). Pearson's correlation coefficient was used to derive the correlation between the evaluation scores obtained from crowdsourcing and the estimation scores obtained from the model. The MAE shows the average of the absolute errors that express the ambience of the same area. We built two learning models: one, trained on landscape images of Nara, and the other, trained on landscape images of Kyoto. The results are shown in Table 1, which indicate a positive correlation in the case of Nara, and a weak positive correlation in the case of Kyoto. Although, MAE values are not too large, the results suggest that there is room for improvement in the model. However, we observed that machine learning models can quantify and estimate the ambience of an area. Furthermore, we discussed certain insights from these experimental results in Section 4.1.

3.2 Routing

The proposed method can be applied to routing for route recommendation. Our routing method refers to the method by Quercia et al. [1].

$$weight_i = \frac{k_i - \min\{k_i\} \forall_i}{\max\{k_i\} \forall_i - \min\{k_i\} \forall_i}, \quad (1)$$

where $score_i$ indicates the ambience score of the i -th street. Here, k is the cube of the difference between the constant ($maxscore^3$) and the score ($score_i$) to reverse the score. The $maxscore$ is four,

¹<https://developers.google.com/maps/documentation/streetview/overview>

²<https://crowdsourcing.yahoo.co.jp/>

³<http://places2.csail.mit.edu/index.html>

Figure 2: Shortest (a) and retro (b) routes from the starting point (34.6796, 135.8260) to destination (34.6749, 135.8306), and examples of landscapes seen in each route in Nara

which is the maximum value of the evaluation in crowdsourcing. In other words, $maxscore^3$ represents the maximum possible value of $score_i^3$ (64). The weights were calculated by normalizing the maximum score value to zero and the minimum value to one. The smaller the weight is, the more retro the street is. Next, the weights calculated using Eqs. 1 were assigned to each street in the area. We further performed routing using Dijkstra’s algorithm [4], which derives the route with the lowest acquisition cost.

We demonstrated how our model can function in a real-world setting in the city of Nara. Given a starting point and destination, Figure 2(a) shows the shortest route as calculated through the Google Maps route search function, while Figure 2(b) shows the route generated using the proposed method. P1 and P2 are the points that are passed in each route. The landscape at each point is shown in Figure 2. As can be observed, the shortest route (a) passes P1 through a featureless residential area, while the retro route (b) passes P2, where a temple is situated near the residential area. Thus, the proposed method succeeded in suggesting a new and interesting route.

4 DISCUSSIONS

4.1 Analysis of Samples

In this section, we compare the evaluation results for each experiment. We determined that the streets or cells with the smallest errors featured famous retro sites, such as Gango-ji Temple (Nara) and the Kyoto Imperial Palace (Kyoto). This suggests that visual distinctiveness and cultural prominence yield more accurate deep learning estimations and more decisive crowdsourced responses. For instance, we can observe the landscape images of the old wooden lattice works in Nara, the lush greenery of the Kyoto Imperial Palace, and the clear castle walls of Nijo Castle in Kyoto.

Figure 3: Comparison of crowd-sourced (a) and estimated (b) heat maps in Kyoto: P3 was commonly perceived as retro, but P4 was not, and examples of landscapes seen in Kyoto

Meanwhile, crowd-sourced values for areas without landmarks (including residential areas) tended to be inconsistent, while the associated deep learning estimations demonstrated the tendency to be ambiguous. As an example, Figure 3 compares the crowd-sourced and estimated heat maps from the Kyoto case study. P3 indicates the area in front of the famous castle square. The nearby stone walls and lush greenery contribute to the appeal of this area, and it is deemed to be retro by both, the crowd-sourced and estimation results. However, P4 is a small street with no landmarks nearby. The parking lots and residential areas leave little impression on people. Thus, both results indicate that this area is not retro. The experiment also suggests that people’s impressions of a part of the area without landmarks are inconsistent, and it is possible that subjective information may not be able to be captured correctly through crowdsourcing. We believe that the regression model improves the accuracy of crowdsourced results by standardizing subjective responses. It may also be important to consider crowdsourcing methods that do not depend on the motivations or attributes of crowd workers.

Table 2: Estimation performance of models trained on different city landscape images

Model	Estimation	Correlation	MAE
Kyoto	Nara	0.125	0.836
Nara	Kyoto	-0.129	0.816

4.2 Possible Improvements of the Model

Our experiments proved the possibility of computationally estimating the ambience of a location based on landscape images; however, the results obtained are not sufficiently accurate. This may be partly, because we used significantly less training data than previous studies [6], which used the same pre-trained model. Consequently, there is a high possibility of unexpected over-fitting. To circumvent this issue, future studies must increase the training data through data augmentation [7] and similar methods.

Another possible approach is to increase the number of image collection points in Google Street View, or to supplement these with landscape images acquired from other media. The pre-trained GoogLeNet model used in this study achieved outstanding image recognition by means of fine-tuning and transfer learning [8, 9]. It's implementation has hence, brought on remarkable effects in fields, such as medicine [10, 11]. Based on this precedent, we will also investigate the use of transfer learning to estimate the ambience of a street.

4.3 Limitations

We discussed the generalizability of the proposed model across cities. Specifically, we conducted additional experiments to verify whether the model trained on Nara landscape images could be used with Kyoto landscape images, and vice versa. Table 2 shows the results of these experiments, which are considerably worse than those in Table 1. This indicates that the models are not sufficiently versatile to work with images from other cities. This may be because the subjective retro judgments of people vary from region to region. For example, in Nara, people are more likely to describe old private houses and wooden architecture as “retro,” while people in Kyoto tend to reserve this description for places designated as tourist attractions, or places that are rich in greenery, such as the Imperial Palace and the castle. As a result, it is difficult for a model trained on the landscape images of one city to properly estimate the ambience of the landscape of another city.

Finally, we considered the generalizability of ambience indicators other than “retro.” Because the proposed method for measuring ambience relies on crowd-sourced labels and landscape image features, it lends itself well to visual indicators, such as “lush” and “beautiful.” One caveat, however, is that these indicators are often confused by people. However, it is more difficult to apply the proposed method to indicators involving human movement or auditory features, such as “noisy,” “quiet,” and “lively.” To apply the method to such indicators, it may be necessary to supplement the landscape images with crowd-sourced information regarding time and season.

5 CONCLUSION

This study employed crowd-sourced labels, landscape images, and deep learning to estimate subjective information about geographic environments that cannot be obtained through conventional sensors. As a case study, we chose “retro” as an ambience indicator and Nara and Kyoto as target areas. First, we surveyed people's impressions of these areas through crowdsourcing, calculated the ambience score, and thereby quantified the ambience. Next, we used a deep learning model to extract notable features from the landscape images and applied linear regression to estimate the ambience score of unknown landscape images based on the extracted features. By applying this method, we compared the scores obtained through crowdsourcing, with those estimated by the model. We established a positive correlation between the two scores with an error that was not too large. The model successfully estimated the ambience with reasonably high accuracy when there were famous sites or tourist attractions in the area. This method facilitates creative route recommendations. For example, a tourist could select their route based on ambience attributes such as retro, rather than length or duration, thereby customizing their journey to the destination in a tourist attraction. With further improvements in crowdsourcing and model performance, the model will prove useful for quantifying and estimating the subjective ambience of local areas.

ACKNOWLEDGMENTS

Part of this work was supported by a JSPS Grant-in-Aid for Scientific Research JP19K20279.

REFERENCES

- [1] Daniele Quercia, Rossano Schifanella and Luca Maria Aiello, “The Shortest Path to Happiness: Recommending Beautiful, Quiet, and Happy Routes in the City,” In Proc. of HT 2014, pp. 116–125, 2014
- [2] Shoko Wakamiya, Panote Siriaraaya, Yihong Zhang, Yukiko Kawai, Eiji Aramaki and Adam Jatowt, “Pleasant Route Suggestion based on Color and Object Rates,” In Proc. of WSDM 2019, pp. 786–789, 2019
- [3] Akio Muranaka, Tomoki Nakaya, “Geographic Variation in Tourism Demand by Historical Landscape Conservation in Terms of the Contingent Behavior of Potential Tourists to Kyoto,” Annals of the Association of Economic Geographers, 2012, Vol. 58, Issue 4, pp. 336–356, Released May 19, 2017
- [4] Edsger W. Dijkstra, “A Note on Two Problems in Connexion with Graphs,” Numerische Mathematik, 1, 269–271, 1959.
- [5] Christian Szegedy, Wei Liu, Yangqing Jia, Pierre Sermanet, Scott Reed, Dragomir Anguelov, Dumitru Erhan, Vincent Vanhoucke, Andrew Rabinovich, “Going Deeper With Convolutions,” In Proc. of CVPR, pp. 1–9, 2015
- [6] Carl Doersch, Saurabh Singh, Abhinav Gupta, Josef Sivic, and Alexei A. Efros, “What Makes Paris Look like Paris?,” ACM Transactions on Graphics (SIGGRAPH 2012), Vol. 31, No. 3, 2012
- [7] Connor Shorten and Taghi M. Khoshgoftaar, “A Survey on Image Data Augmentation for Deep Learning,” J Big Data 6(60), 2019
- [8] Relja Arandjelović, Petr Gronat, Akihiko Torii, Tomas Pajdla and Josef Sivic, “NetVLAD: CNN Architecture for Weakly Supervised Place Recognition,” IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 40, no. 6, pp. 1437–1451, 2018
- [9] Hoo-Chang Shin, Holger R. Roth, Mingchen Gao, Le Lu, Ziyue Xu, Isabella Noguees, Jianhua Yao, Daniel Mollura and Ronald M. Summers, “Deep Convolutional Neural Networks for Computer-Aided Detection: CNN Architectures, Dataset Characteristics and Transfer Learning,” IEEE Transactions on Medical Imaging, Vol. 35, no. 5, pp. 1285–1298, 2016
- [10] Tiantian Fang, “A Novel Computer-Aided Lung Cancer Detection Method Based on Transfer Learning from GoogLeNet and Median Intensity Projections,” In Proc. of CCET, pp. 286–290, 2018
- [11] Mostafa Mehdipour Ghazi, Berrin Yanikoglu and Erchan Aptoula, “Plant Identification Using Deep Neural Networks via Optimization of Transfer Learning Parameters,” Neurocomputing, Vol. 235, pp. 228–235, 2017