Sightseeing Value Estimation by Analyzing Geosocial Images

Yizhu Shen

Department of Social Informatics, Graduate School of Informatics,
Kyoto University, Kyoto, Japan
E-mail: shen@db.soc.i.kyoto-u.ac.jp

Min Ge

Department of Social Informatics, Graduate School of Informatics,
Kyoto University, Kyoto, Japan
E-mail: gemin@db.soc.i.kyoto-u.ac.jp

Chenyi Zhuang

Department of Social Informatics, Graduate School of Informatics,
Kyoto University, Kyoto, Japan
E-mail: zhuang@db.soc.i.kyoto-u.ac.jp

Qiang Ma

Department of Social Informatics, Graduate School of Informatics,
Kyoto University, Kyoto, Japan
E-mail: qiang@i.kyoto-u.ac.jp

Abstract: Recommendation of points of interests (POIs) is drawing more attention to meet the growing demands of tourists. Thus, a POI's quality (sightseeing value) needs to be estimated. In contrast to conventional studies that rank POIs on the basis of user behavior analysis, this paper presents methods to estimate quality by analyzing geo-social images. Our approach estimates the sightseeing value from two aspects: (1) nature value and (2) culture value. For the nature value, we extract image features that are related to favorable human perception to verify whether a POI would satisfy tourists in terms of environmental psychology. Three criteria are defined accordingly: coherence, image-ability, and visual-scale. For the culture value, we recognize the main cultural element (i.e., architecture) included in a POI. In the experiments, we applied our methods to real POIs and found that our approach assessed sightseeing value effectively.

Keywords: Points of Interests; Sightseeing value; Geosocial image; human perception; image processing; UCG Mining.

Reference to this paper should be made as follows: Y. Shen, M. Ge, C. Zhuang and Q. Ma (2016) 'Sightseeing Value Estimation by Analyzing Geosocial Images', *International Journal of Big Data Intelligence*, Vol. x, No. x, pp.xxx–xxx.

Biographical notes: Yizhu Shen is a masters student in the Department of Social Informatics at Kyoto University. Her research interests include information retrieval, multimedia information processing, and Web mining.

Min Ge is a masters student in the Department of Social Informatics at Kyoto University. His research interests include Web mining and user behavior analysis.

Chenyi Zhuang is a PhD candidate at the Graduate School of Informatics, Kyoto University, where he is also serving as a research fellow in the Japan Society for the Promotion of Science (JSPS). He received a BS degree in SE from Nanjing University in 2011 and an MS degree in informatics from Kyoto University in 2014. In between, he participated in the Asia Future Leaders Program funded by the Bai Xian Education Foundation. Starting in the fall of 2015, he worked as a visiting researcher in the Knowledge Mining Group at Microsoft Research Asia. His current research primarily involves social multimedia mining, social network analysis, and urban computing.

Qiang Ma received his Ph.D. degree from the Department of Social Informatics, Graduate School of Informatics, Kyoto University in 2004. He was a research fellow (DC2) of the JSPS from 2003 to 2004. In 2004, he joined the National Institute of Information and Communications Technology as a research fellow. From 2006 to 2007, he served as an assistant manager at NEC. From October 2007, he joined Kyoto University and has been an associate professor there since August 2010. His general research interests are in the area of databases and information retrieval. His current interests include Web information systems, Web mining, knowledge discovery, and multimedia mining and searching.

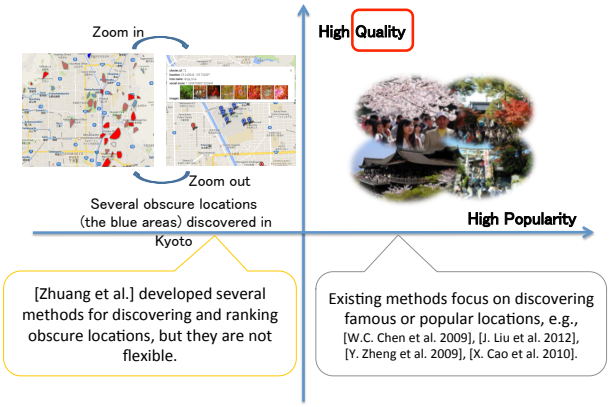


Figure 1: Two dimensions to describe POIs proposed by Zhuang et al. (2014).

1 Introduction

Nowadays, travel plays a larger part of people's lives. Benefiting from Social Networking Services (SNSs) and advances in mobile devices, people can share their experiences on the Internet during travel. The vital information this sharing contains provides researchers with excellent opportunities for discovering and ranking points of interests (POIs). For instance, Zheng and Xie (2001) treated GPS traces, Chen et al. (2009) images, Liu et al. (2012) check-ins, and Hasegawa et al. (2012) tweets as different kinds of user votes to help gather tourism knowledge. To evaluate these votes, a method is needed to evaluate the quality of a POI.

Although many researchers, such as Zheng and Xie (2001) and Liu et al. (2012), have done work on POI recommendation, much is still unexplored. Based on a survey by Zheng et al. (2011), the growing geo-referenced and community-contributed media resources have generated huge amounts of detailed location and event tags, covering not only popular landmarks but also obscure ones. As shown in Figure 1, we can divide POIs into four quadrants on the basis of two dimensions: quality and popularity (Zhuang et al. (2014)).

Located in the quadrant with high sightseeing quality but low popularity, an obscure sightseeing location can be a interesting choice for in-depth travel to not only enjoy beautiful scenery but also experience local culture, especially for repeat tourists who have already visited the most famous places in an area. In some senses, such locations may be potentially valuable sightseeing resources that need to be developed and promoted. However, because obscure locations almost never have enough visits or votes on the Internet, the conventional authority based analysis used to recommend popular POIs is not useful. Zhuang et al. (2015) and Zhuang et al. (2014) presented methods to discover and rank obscure locations. However, their methods still rely on analyzing few users' behaviors and the type of scenery objects (cherry blossom and maples are used as examples in their work), which make their solutions inflexible.

In this paper, by analyzing geo-social images, we present a general approach to estimate the quality of both popular and obscure sightseeing spots. When people experience a landscape, information is derived through senses, organized, and interpreted by human perception (Kaplan (1978)). In this way, a mental model (Bourassa (1991)) has been devised in which human perception is affected by three aspects:

1. biological factors according with evolutionary theory,
2. cultural factors depending on cultural background, and
3. individual factors resulting from individual differences in personality traits.

In accordance with this mental model, cultural factors vary among peoples, individual factors vary from person to person, and biological factors can be treated as cross-cultural commonalities for human perception of landscapes. Therefore, we focus on the criteria served by the biological factors, which interpret the landscape from a physical level to a psychological level. By introducing the criteria (i.e., coherence, complexity, disturbance, stewardship, image-ability, visual-scale, naturals, historicity, and ephemera, defined in environmental psychology by Tveit et al. (2006)), we calculate image features as indicators to estimate quality. Because these criteria are interrelated and interact, our approach mainly focuses on four key criteria: coherence, imageability, visual-scale, and historicity. The first three are related to nature value (NV) (i.e., sightseeing quality estimation from an environmental psychological perspective), while the fourth views sightseeing spots from the angle of culture value (CV) (i.e., the sightseeing value from a cultural perspective).

Instead of discovering well-known or obscure spots, our work focuses on ranking the spots on the basis of their nature and culture values. It is to say, our methods can analyze and rank well-known spots and obscure spots. To the best of our knowledge, this is the first attempt to estimate sightseeing value by utilizing environmental psychology. To summarize, we make the following major contributions:

- Content based methods for estimating sightseeing spots from a nature aspect: By introducing the qualitative nature criteria defined in environmental psychology, we quantize three (i.e., *coherence, image-ability and visual-scale*) to estimate a POI's NV. To extract the indicators for the quantization, we devise several new algorithms to calculate the visual features from geo-social images taken of or at the target POI.
- A time-based analysis: Because of seasonal variations, a time series based analysis is further made to obtain dynamic evaluation results for

ranking POI candidates, on the basis of which we can recommend different spots to users on the basis of the season in which they are planning to visit.

- A content based method for estimating sightseeing spots from a culture aspect: Different from the human-based culture factors mentioned previously, here culture refers to the inherent value held by the spot, which means we only estimate culture objectively without considering the cultural backgrounds of various tourists. Since some POIs contain several artificial elements (e.g., architecture), a heuristic method is developed to measure the CV.

2 RELATED WORK

In this section, we first present the conventional related work on ranking POIs. Then, several studies on human perception for landscape environment are introduced followed by related work using image analysis. Lastly, work related to culture value is also discussed.

Ranking POIs In the research into estimating sightseeing quality, Luo et al. (2011) conducted a survey showing that collections of geo-multimedia, which are a result of sightseeing experiences shared among web communities, are widely used in trip recommendations. Ji et al. (2009) modeled the relationships of scene/landmark and scene/authorship as a graph and adopted two popular link analysis methods (PageRank and HITS) to mine representative landmarks. Zheng et al. (2009) aimed to mine interesting locations and regular travel sequences in a given geospatial region on the basis of multiple users' GPS trajectories. They first modeled multiple individuals' location histories with a tree-based hierarchical graph. Then, by using the graph, they developed a HITS-based inference model that infers the interest in a location. Zheng and Xie (2001) further developed a recommendation system. Liu et al. (2012) presented a joint authority analysis framework to discover areas of interest with geo-tagged images and check-ins instead of GPS traces. Hasegawa et al. (2012) attempted to organize travel related tweets by considering the spatio-temporal continuity of user-behaviors during travel. By merging such fragmented tweets, users' travel experiences can be detected.

In these studies, GPS traces, images, check-ins, and tweets are treated as different kinds of user votes to help gather tourism knowledge. Authority based analysis, like "rank-by-count" and "rank-by-frequency" in a vote manner, is the basis for most of this trip recommendation research. However, for an obscure location, not enough visits or votes on the Internet are generated. Thus, conventional authority based analysis used to recommend popular sightseeing locations is not suitable. Therefore, in our research, human perception is introduced as a solution.

Human perception There have been many systematic analyses and studies on the human perception of landscape environments. Hartig (1993) suggested that the settlement in a landscape mainly resulted from evolutionary, sociocultural, and motivational forces. Differences in natural landscape preference between user groups coming from different backgrounds is proved in experiments done by Berg et al. (1998). Ohta (2001) proposed 11 cognitive criteria for evaluating natural landscapes and summarized a qualitative common structure for natural landscape cognition. Tveit et al. (2006) developed an abstract framework for people's interpretation for a landscape from concept level to indicator level. In this framework, they proposed nine concepts for landscapes.

This previous work presented the concepts and design disciplines for sightseeing value assessments and landscape restorations. In contrast, we present a novel quantitative analysis method for assessing landscapes by exploiting geo-social images. To the best of our knowledge, our work is the first attempt at quantitatively estimating sightseeing values from natural and cultural perspectives.

Nature value In the image-processing field, researchers are trying to discover the relationships between images and human perception. Estimating the aesthetic quality of a photo is highly related to our work. Tang et al. (2013) extracted both regional and global high-level features and tried to build connections between photo qualities and technical rules shared by photographers. Datta et al. (2006) described the aesthetic quality by selecting low-level features on the basis of artistic intuition. Furthermore, more real-scene dependent features such as sky illumination (Dhar et al. (2011)) and landscape types (Yin et al. (2012)) have also been considered for improving quality assessment.

In addition to this related work focusing on the evaluation for a single image rather than a real scene, Berman et al. (2014) have produced research that is quiet similar to ours: they tried to uncover low-level image features related to human perception of naturalness. Furthermore, Hunter and Askarinejad (2015) summarized the properties predicted to be important in usual environmental theories and listed some measurable corresponding physical attributes for landscape preference.

However, we argue that the low-level features implemented in this related research, such as color and spatial properties, are insufficient. To solve our problem, such features must be leveraged to a higher level.

Culture value Culture is a very abstract concept including many sub-concepts such as art, design, and history and varies among countries and regions. For example, famous elements of Japanese culture include Niwa (traditional Japanese architecture), Ikebana (flowers arrangement), Bonsai (trees grown in containers), Katana (Japanese sword), and Kimono

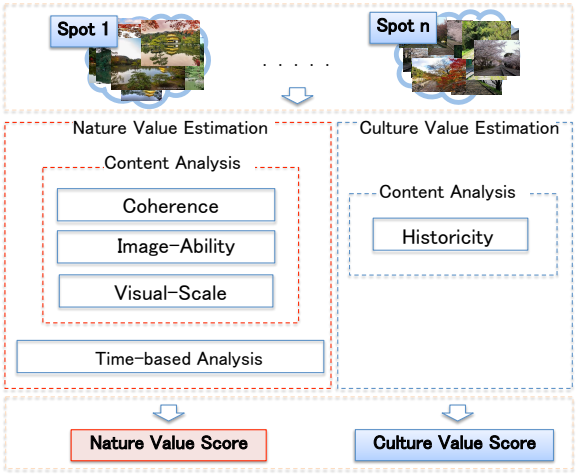


Figure 2: Overview of our approach.

(traditional Japanese female costume) Mente (2011). In this paper, the purpose of our method is to evaluate the CV of a sightseeing spot using images. However, only a few of these cultural elements frequently appear in the images taken by tourists. The most common cultural element found in the images is traditional architecture, which plays a very import role in culture evaluation in the method proposed by Emmons et al. (2012). Traditional costume is another cultural element that varies greatly among cultures (Harrold et al.) and contributes a lot to culture evaluation (Pendergast et al. (2003)).

Architecture parsing has been studied by Berg et al. (2007), detecting by Toshev et al. (2010), style classification by Xu et al. (2014), and clothes style classification by Bossard et al. (2012). However, as far as we know, no work has combined the evaluation of sightseeing spot's cultural value and the detection of these cultural elements. In this paper, we build a bridge connecting these two areas.

3 METHODOLOGY

In this section, we introduce our methods to estimate sightseeing values of a spot from natural and cultural perspectives. As shown in Figure 2, the input data of our methods are spots with geo-tagged images, and the output data are two sightseeing values from natural and cultural perspectives. For a preliminary process, we can obtain sightseeing spots by applying clustering methods, such as DBSCAN, to the geo-tagged images.

3.1 NV Evaluation

According to landscape perception theory, the quality of a landscape is affected by multiple factors. Therefore, first, we design a corresponding-image analysis method for each factor per each image. Then, we integrate these factors to obtain the NV of a spot using a set of images.

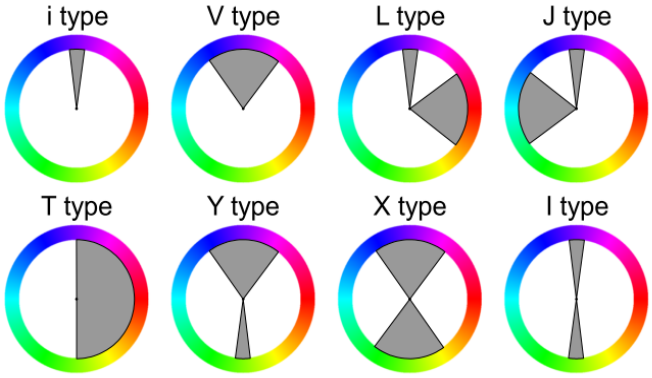


Figure 3: Harmonious hue templates by Matsuda (1995).

In the final step, we arrange all the scores in a time series way, by which the seasonal issues are considered.

Based on the study of environmental psychology by Tveit et al. (2006), nine criteria should be considered for landscape assessments: coherence, complexity, disturbance, stewardship, image-ability, visual-scale, naturals, historicity, and ephemera. To estimate the NV of a given spot using images, we focus on coherence, image-ability and visual-scale, which are more realizable by utilizing image processing methods on the basis of previous research.

3.1.1 Coherence

The coherence relates to the unity of a scene, enhanced by the degree of repetition of color and texture patterns Tveit et al. (2006). On the basis of this definition, we consider color harmony and repeated patterns as detailed indicators for estimating the coherence of spots.

Color Harmony. Intuitively, colorful landscapes are worth visiting. In this sense, we introduce color harmony as an indicator to estimate the NV on the basis of coherence. Matsuda (1995) proposed eight harmonious hue templates defined in a HSV space. As shown in Figure 3, each harmonious hue template contains a gray sector, which is the harmonic hue distributor for an image. All the areas and relative position relationships of sectors are fixed and only the rotation angle may change. An image that has a hue distribution fitting one of these templates can be regarded as having high color harmony.

Given an image, we use a harmony distance to calculate the difference between an image's original hue distribution from the harmonious hue templates. The harmony distance with the most suitable template is defined as the color harmony score. We define each harmonious hue template T_m as:

$$T_m = \{(a_m, \omega_{m,k}); k = 1, \ldots, K_m\} \quad (1)$$

$m \in$ {i, V, L, J, T, Y, X, I} means the eight templates shown in Figure 3. The notation $K_m \in \{1, 2\}$ is the number of sectors in the m-th template and $\omega_{m,k}$ is the

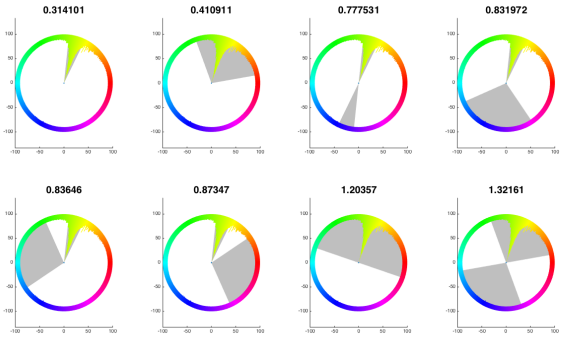


Figure 4: Harmony distance calculated against each template for given image. Top-left is most matched template while bottom-right is worst.

area of the k-th sector in the m-th template. a_m is the rotation angle for the template m. The harmony distance from a given hue distribution to the m-th template is calculated by an appropriate a_m, which is introduced to minimize the distance as follows.

$$a_m \sum_h M(h) L_m(h, m) \tag{2}$$

where $h \in \{0, ..., 359\}$ is the index on each hue template. M is a normalized hue distribution for an image and $L_m(h, m)$ is the loss function for T_m in the hue position h. To define the loss function $L_m(h, m)$, we first introduce a Gaussian distribution $D(h, a_m, \omega_{m,k})$, which is used to adjust the penalty of the loss function. The closer an index h approaches the boundaries of the sector k in the template m, the larger the penalty will be.

$$D(h, a_m, \omega_{m,k}) = \frac{1}{\sqrt{\pi \omega_{m,k}}} \exp(-\frac{2|h - a_m|^2}{\omega_{m,k}^2}) \tag{3}$$

$$ch(i) = L_m(h, a_m) = k(D(h, a_m, \omega_{m,k}))|k| + 1$$

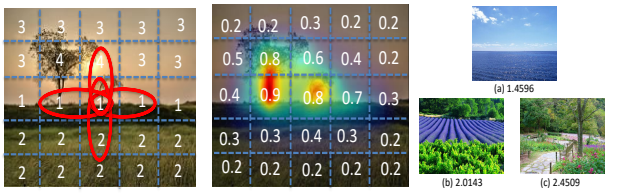
when $\forall k \in \{1 \dots K_m\}, |h - a_m| \geq \frac{\omega_{m,k}}{2}$ (i.e., h is in the sector k); and

$$= \frac{\omega_{m,k}}{2\pi^2} \sum_{|h^* - a_m| \geq \frac{\omega_{m,k}}{2}} (D(h^*, a_m, \omega_{m,k}) + 1) + \sum_{k_i \in \{1,\dots,K_m\} - k} (D(h, a_m, \omega_{m,k_i}) + 1)$$

when $\exists k \in \{1, \dots, K_m\}, |h - a_m| < \frac{\omega_{m,k}}{2}$ (i.e., h is out of the sector k).

Figure 4 shows our calculation results using this algorithm. The template with the lowest harmony distance is considered as the most matched one for the given image, and the distance value is used as the color harmony based score.

Repeated Pattern. If a landscape contains blocks with the same or repeated patterns, the scenery is ordered and its coherence is considerably high. We consider a repeated pattern as repeated or similar blocks shown in an image. In advance, we divide an image into blocks of 15×15 pixels and represent each block using a HSV space-based histogram. We apply the Self Organizing Map (SOM) of Kohonen and Somervuo (1998) to cluster these blocks into 16 (4×4) groups.



(a) $n_{i,j}$ (b) Saliency score (c) Calculated scores

Figure 5: Examples of calculating repeated pattern based score.

To reveal the relative position of blocks, for any two groups i and j, we use $n_{i,j}$ (see Figure 5a) to denote the number of times group j is adjacent to group i. The normalized $n_{i,j}$ can be seen as an occurrence probability of such a case.

Since all the blocks hold different saliency for perception, we calculate the average saliency score $a_{i,j}$ for groups i and j by using the saliency map method developed by Harel et al. (2006). An example is shown in Figure 5b. On the basis of the idea of weighted entropy provided by Guia (1971), we obtain the repeated pattern score $rp(i)$ for image i by the following formula. Figure 5c presents three examples, by which we can observe that [the more similar and ordered the blocks in an image, the lower its repeated pattern score. A low repeated pattern score means high coherence.

$$rp(i) = -\sum_{i=1}^{15} \sum_{j=i}^{15} a_{ij} n_{ij} \log(n_{ij}) \tag{4}$$

We annotate $co(i)$ as the coherence for image i, and the $CO(s)$ as the coherence for spot s.

$$co(i) = \frac{ch(i) + rp(i)}{2} \tag{5}$$

$$CO(s) = \frac{1}{n} \sum_{j=1}^{n} co(i) \;\; i \in \{i_{s1}..i_{sn}\} \tag{6}$$

3.1.2 Image-ability

The image-ability, which is defined as a strong visual image created by the landscape that gives people a distinguishable and memorable experience, is conceptionally similar to the photo quality assessment developed by Tveit et al. (2006). Therefore, we exploit photo quality assessment methods to estimate the image-ability of a sightseeing spot. The idea is simple: if a spot has photos with high image-ability, its sightseeing quality is reasonably high.

We use a machine learning method for this task. The database used for training contains the images categorized as landscapes in the aesthetic visual analysis (AVA) dataset of Murray et al. (2012), which contains 250,000 images with aesthetic scores and semantic labels. We sort the images by their average scores and prorate the scores with a value ranging from 1 to 5. Inspired by the work done by Tang et al. (2013), Dhar et al. (2011), and Yin et al. (2012), we extract three low-level features to describe the whole image: the histogram of oriented gradients (HOG) (Dalal and Triggs (2005)), color moment (mean and standard deviation for RGB channel) (Stricker and Orengo (1995)), and local binary patterns (LBP) (Ojala et al. (1996)). HOG is widely used for object detection. LBP is found to be effective for texture classification and color moments, which characterize color distribution, and is often used in image classification. For the training model, since a comparable output is expected, we use the cluster-weighted modeling (CWM) developed by Ojala et al. (1996) to do the regression and use the predicted value as the image-ability score, where the value range is from 1.0 to 6.0.

We annotate $im(i)$ and $IM(s)$ as the image-abilities for an image i and spot s, respectively.

$$IM(s) = \frac{1}{n}\sum_{j=1}^{n} im(i) \;\; i \in \{i_{s1}..i_{sn}\} \tag{7}$$

3.1.3 Visual-scale

The visual-scale is defined as a perceptual unit that reflects the experience of landscape rooms, visibility, and openness (Tveit et al. (2006)). To calculate this criteria, we use the GIST based method introduced by Oliva and Torralba (2001) to estimate the openness and depth using an image. The value range of both openness and depth is from 1 to 6. Here, openness refers to the view-shed size or the degree of occlusion of a landscape. The depth is more relevant to the max visual distance. Since both openness and depth indicate the visual-scale of a landscape, we calculate these two values $op(i)$ and $dp(i)$ by using the model provided by Oliva and Torralba (2001) and use the average to calculate the visual-scale score for a spot.

We annotate $vi(i)$ and the $VI(s)$ as the visual-scale for an image i and spot s, respectively.

$$vi(i) = \frac{op(i) + dp(i)}{2} \tag{8}$$

$$VI(s) = \frac{1}{n}\sum_{j=1}^{n} vi(i) \;\; i \in \{i_{s1}..i_{sn}\} \tag{9}$$

3.1.4 NV Calculation

We denote the input spot set as $S = \{s_1...s_n\}$. Each spot s_i is represented by an image set. Because the NV of a spot varies by season, we divide the images into 12 months and try to implement these three evaluation method dynamically. First, for all defined criteria (i.e., coherence, image-ability, and visual scale), we construct three corresponding matrices: M^c, M^i, and M^v. Hereinafter, M is denoted as one of the three matrices. $M_{i,j}$ is the average score of the target criteria for spot s_i in month $j = \{1...12\}$. Then, on the basis of the M, three aspects are considered to evaluate s_i: overall level, durability, and uniqueness. The overall level and durability are used to assign a high value for a spot with high and stable nature perception, which is perceived as a sightseeing spot suitable for a large number of tourists. Besides, since people tend to make more effort to find something special, we assign uniqueness a higher value while the other spots have relatively low values for each month.

(1) Overall level $Avg(s_i)$ of spot s_i.

$$Avg(s_i) = \frac{1}{12}\sum_{j=1}^{12} M_{i,j}; \;\; i \in \{1..|S|\} \tag{10}$$

(2) Durability $Dub(s_i)$ of spot s_i.

$$Dub(s_i) = \sqrt{\frac{1}{12}\sum_{j=1}^{12}(M_{i,j} - \frac{1}{12}\sum_{j=1}^{12} M_{i,j})^2}, \;\; i \in \{1..|S|\} \tag{11}$$

(3) Uniqueness $Uni(s_i)$ of s_i.

$$Uni(s_i) = \frac{1}{12}\sum_{j=1}^{12} f(M_{i,j}) \;\; i \in \{1..|S|\} \tag{12}$$

$$f(M_{i,j}) = max\{0, M_{i,j} - \frac{1}{|S|}\sum_{i=1}^{|S|} M_{i,j}\} \tag{13}$$

Finally, the coherence, image-ability, and visual-scale based NV scores are calculated by their respective means.

$$NV(s_i) = \frac{1}{3}(Avg(s_i) + Dub(s_i) + Uni(s_i)) \tag{14}$$

3.2 CV Evaluation

The purpose of this part is to estimate the cultural value of sightseeing spots on the basis of images taken by tourists. There are two challenges. The first is that culture is a very abstract concept and hard to estimate. Our solution is to decompose a sightseeing spot into several objects and estimate the cultural value of each object. The second challenge is that our estimation is wholly based on images, which means we have to choose the objects that appear commonly in images taken by tourists. We summarize five cultural elements that obviously affect the cultural value and commonly appear in photos. Table 1 shows the relationships among them. Architecture, its adornment, and traditional costumes are very important cultural elements, as mentioned in Section 2. In addition to these, color preference is another vital part of culture according to Hochman and

Table 1 Relationships among cultural elements

	Object-dependent	Object-independent
Static	Architecture, Adornment	Color Preference
Dynamic	Traditional, Costume	Activity, Event

Schwartz. For example, people who have taken photos in New York seem to prefer blue-gray, while people in Tokyo like red and yellow more. Besides, festivals and some cultural events reproduce scenes from traditional culture. If a cultural element does not change over the year, we say it is static and dynamic otherwise. If a cultural element can be defined on the basis of only one object, we say it is object-dependent and object-independent otherwise. For example, color preference does not change over the year, but we cannot define the color preference by only one object, so it is static and object-independent. Conversely, traditional costume is worn by people able to go to any sightseeing spots they like, so it is object-dependent and dynamic. In this paper, we only choose one of cultural element (architectural style, which at the same time includes color preference) to estimate the cultural value of sightseeing spots.

Styles of architecture vary greatly among countries and regions. Traditional architecture increases the cultural value of sightseeing spots. Therefore, the aim of this part is to detect architectural objects and classify them into different architectural style. One sightseeing spot can have many different kinds of architecture in one sightseeing spot such as towers, temples, and bridges. One architecture category may contain many objects. We assume that the cultural value of sightseeing spots is positively related with the cultural value of each architectural object located in the landscape. Therefore, Figure 6 shows three steps for evaluating cultural value for each architectural object: category clustering, architectural object separation, and cultural value scoring.

3.2.1 Category Clustering

A set of images of a sightseeing spot contains different kinds of objects such as architecture, natural scenes, and tourists. The purpose of this step is to gather images belonging to the same category together. VGG net provided by Simonyan and Zisserman (2014) is a very well-known convolutional neural network that classifies images with outstanding accuracy. In addition to good classification performance, VGG net also generates features of high quality. Here, we use the output of the last fully connected layer, which contains 4096 dimensions as feature descriptors, and cluster them by a k-means clustering method.

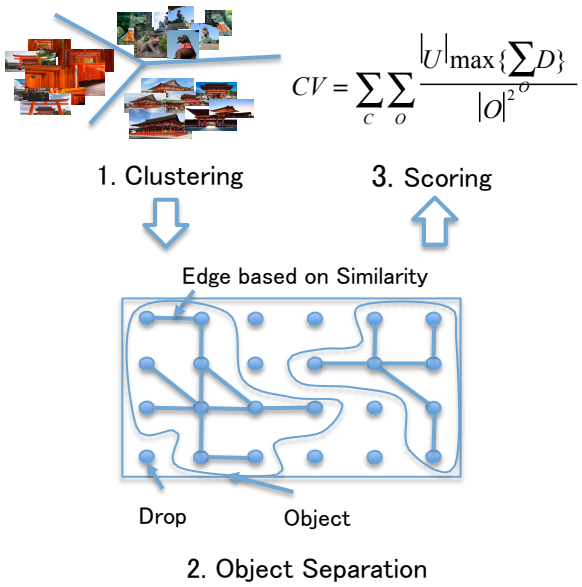


Figure 6: Three steps for cultural value evaluation.

3.2.2 Architectural Object Separation

After clustering, each cluster will contain multiple objects. The images depicting the same object are similar but obviously different from the images depicting other objects. SIFT provided by Lowe (2004) is a highly suitable descriptor for this task. We define the distance between images A and B by the following formula.

$$D(A,B) = \frac{(n_A + n_B)\sum_{M_{AB}} score_{AB}}{2|M_{AB}|^2} \tag{15}$$

where n_A and n_B denote the number of SIFT points found in images A and B respectively, M_{AB} is a set of matching points, and $score_{AB}$ is a set of matching scores. Different size and resolution images will generate different numbers of SIFT points. Therefore, we first calculate the average points found in images A and B. If images A and B depict the same object, we can find a large number of matching points with low scores. In other words, the distance between images A and B is negatively related with the size of set M_{AB} and positively related with the score of each matching point. Therefore, the sum of scores of matching points is the numerator, and the size of set M_{AB} is the denominator. Finally, the distance formula is multiplied by the average SIFT points in images A and B to eliminate the effect of image size and resolution.

For images in the same cluster, if the distance between two images is smaller than a threshold, we assume that there is an edge between them. Each image is assumed to be a vertex. Therefore, we obtain an image graph for each cluster. An object is defined as a connected subgraph of each image graph. Images in a connected subgraph will be treated as depicting the same object because they are similar enough.

3.2.3 CV Calculation

In this step, we train a Deformable Part Model (DPM) provided by Felzenszwalb et al. (2010) to classify architectural objects and give a CV score. The training set of the DPM consists of several examples of famous architecture widely regarded as of high culture quality. The DPM gives each image a score of classification confidence. We assume that the score of confidence is positively related with culture quality. The trained DPM model is conducted on each image of each object. If the max DPM score of an object is smaller than a threshold, we regard this object as unrelated with architecture of high culture quality and omit it. The final CV score $CV(s)$ of a sightseeing spot is given by the following formula.

$$CV(s) = \sum_{C} \sum_{O} \frac{|U| \max(\sum_{|O|} D)}{|O|^2} \tag{16}$$

where C denotes the set of clusters, O is the set of objects found in a cluster, and U is a set of users who take the image of the object. D is a confidence vector given by DPM. Intuitively, for the D, the more objects with high DPM scores, the higher the CVS. For the U, if an object has high CV, it should appear in many images taken by different users.

For implementation details, sightseeing spots will be divided into several clusters by using a VGG feature and k-means. Each cluster will contain a number of objects detected by looking for connected subgraphs in an image graph built on the basis of distance defined in step 2.

An object is depicted in several images. The cultural value of a sightseeing spot should be the sum of each architectural object contained in each cluster. The cultural value of a single object is positively related with the confidence score given by the DPM. Here we use the max value of the average confidence score of each image to denote the confidence score of objects. In addition to classification of objects, we also find that user behavior is related with the cultural value of objects. Travelers prefer to photograph objects of high natural or cultural quality. In other words, if an object is of high cultural quality, it should appear in many images taken by different users. Therefore, the number of users who take images of the object is also positively related with cultural value. Here the number of users is divided by the number of images depicting the same object.

Because both the confidence scores given by the DPM and user preference are divided by the number of images, our method evaluates the cultural value of a sightseeing spot while excluding the impact of the number of images. Generally, tourists prefer going to famous sightseeing spots to take images rather than less well-known ones, so we can find more images related to popular spots than obscure spots. Therefore, in our dataset, popular sightseeing spots contain more testing images than others. However, we do not think more images mean higher cultural value. Some obscure sightseeing spots of high culture quality are not crowded with tourists because they have not become widely known. Thus, we have developed this method to evaluate the cultural value of a sightseeing spot regardless of how popular it is.

4 Experiments

In this section, we investigate the effect of criterion calculation methods and demonstrate the performance of our methods for both NV and CV. On the basis of the algorithms introduced in Section 3, for a certain spot, we make an estimation on all the images taken there and use the normalized Discounted Cumulative Gain (nDCG) method by Jarveli and Kekalainen (2002) to evaluate all criterion calculation methods and criteria based evaluation for NV and CV estimation. Furthermore, two baseline methods for NV and one baseline method for CV are compared with our method to demonstrate its effectiveness.

4.1 Dataset

For the experimental data, we initially collected images of 14 sightseeing spots: 7 in Kyoto, Japan and 7 in Suzhou, China. The spots in Kyoto are Fushimi Inari Taisha Shrine, Kinkakuji Temple, Ninnaji Temple, Tenryuji Temple, Shisen-do, Hanami Street, and Kyoto Station. The spots in Suzhou are the Humble Administrators Garden, Tai Lake, Jinji Lake, Tiger Hill, Suzhou Museum, Shantang Street, and Guanqian Street. In this dataset, both high-quality spots that are abundant in natural elements and cultural elements (e.g. Kinkakuji Temple, the Humble Administrators Garden), and low-quality spots that mainly consist of modern architecture (e.g., Kyoto Station) have been considered to promise unbiased experimental data. In our experiment, Shisen-do is not a popular spot (i.e. an obscure spot) but the others are.

We collected about 13,000 geo-tagged images from Flickr for these 14 sightseeing spots. All the images are retrieved by Flickr's keyword based search and verified by their geo-information. For the time-based analysis, we also collected the metadata of images, including the user ID and timestamp. Since we need to extract color features from images in the process of quality calculation, some gray images were removed in advance.

To obtain the ground truth, we employed eight subjects to label each candidate spot with coherence, image-ability, visual scale, nature value, and culture value for all seven spots in Suzhou and seven spots in Kyoto. All subjects were university students from China and Japan. Their different social and national backgrounds gave them different degrees of understanding of target spots. The definitions of each criterion were given to each subject, and subjects could look back and forth at images without any time limit. A five-point scale ranging from "1" for "very low value" to "5" for "very high value" was used, and we regarded the

Table 2 Ground truth: culture and nature scores of each spot

Avg. score	Tennryuji Temple	Ninnaji Temple	Kinkakuji Temple	Shisen-do	Fushimiji Temple	Hanami Street	Kyoto Station	Daigoji Temple[1]
No. of photos	1k	1k	1k	0.4k	1k	1k	1k	1k
Coherence	2.875	2.875	2.75	3.125	3.375	2.5	2.25	2.5
Image-ability	3.125	3.125	3.625	2.875	4.5	3	2.5	4
Visual-scale	3.375	3	3.25	2.625	2.375	2.75	2.375	3.5
Nature value	3.875	3.25	3.25	3.875	2.5	2.625	1.25	4
Culture value	2.875	3.75	4	3.375	4	3.25	2	4.5

Avg. score	Tai Lake	Jinji Lake	Tiger Hill	Suzhou Museum	Humble Garden	Shantang Street	Guanqian Street	Sekizanzenin Temple[1]
No. of photos	1k	0.2k	1k	0.4k	1k	1k	0.3k	0.3k
Coherence	2.625	4.25	3.375	2.25	3.125	2.5	2	3.5
Image-ability	2.875	3.75	3	3.125	3	3.375	2.25	3
Visual-scale	3.75	4.75	3.875	2.375	3.125	2.75	2	2
Nature value	3.875	3.625	4	2	3.75	2.625	1.25	4
Culture value	2.375	2.375	3.125	4.5	3.625	3.75	2.875	3

[1] Obscure spots assessed in our additional experiment.

average of all the subjects labels as the ground truth for a spot. Table 2 shows the label results and the details of our data set.

As an obscure spot, Shisen-do has been included in the dataset. For further investigation of the effect of assessing obscure spots by our methods, we carried out an additional experiment. We added two obscure spots (Sekizanzenin Temple and Daigoji Temple) in our dataset (Table 2), and employed two other subjects [1] to label these spots.

4.2 Evaluation on Nature Value

In accordance with our research, three criteria (coherence, image-ability, and visual-scale) are calculated and used for estimating NV. Hence, first, we evaluate our methods to calculate these criteria.

For each criterion, we calculate the scores for all the images and use the average to describe the target spot. The calculated criteria scores for each spot are normalized to the range of 0 to 1 and compared with corresponding ground truth in Figures 7, 8, and 9. According to the result, despite of the low popularity and small number of images, three obscure spots are calculated in a quite high accuracy, which is the same as

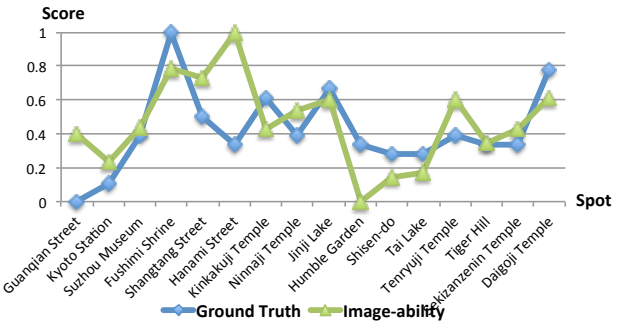


Figure 8: Comparison between image-ability score and ground truth.

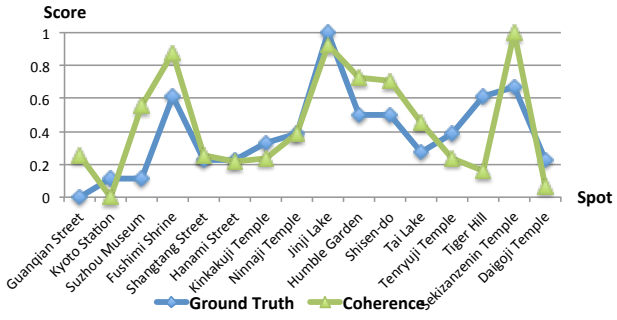


Figure 7: Comparison between coherence score and ground truth.

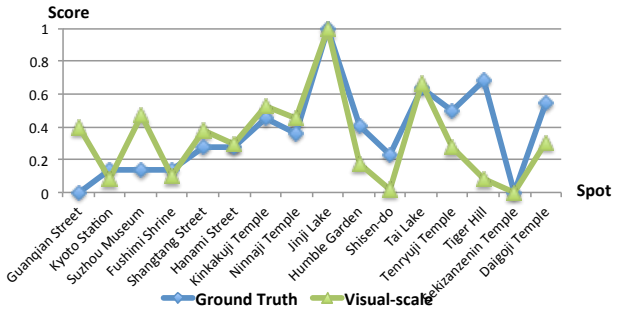


Figure 9: Comparison between visual-scale score and ground truth.

hot spots. Then we use the nDCG method to evaluate all criterion calculation methods with corresponding ground truth, the results of which are shown in Figure 10. It can been that the scores calculated by coherence and visual-scale calculation have a relatively high match rate, while the image-ability calculation deviates from the ground truth.

As introduced in Section 3.1.4, for NV estimation, we calculate an average value for all the images taken in each month by using the time-tag. Then we make a time-based nature evaluation by using these three criterion calculation methods in respective and combined ways and then demonstrate the performance with the nDCG method. Figure 11 shows the evaluation results. The detailed analyses are as follows.

4.2.1 Coherence

Based on our definition, the coherence mainly consists of two aspects: color harmony and repeated pattern.

According to the calculated results for coherence in Figure 9, Fushimi Inari Taish Shrine, Jinji Lake, and Sekizanzenin Temple have relatively high coherences for visual perception. As introduced previously, the coherence is defined as related to the unity of a scene, enhanced by the degree of repeated patterns of color and texture. It is easy to explain that since the images related to Fushimi Inari Taisha Shrine mainly consist of torii (traditional Japanese gates) that are only one color, red, a harmonic color tendency is generated for this spot. For a landside landscape where the scene mostly consists of a clear sky and clear lake, these simple repeated patterns in Jinji Lake gives people a high harmonic

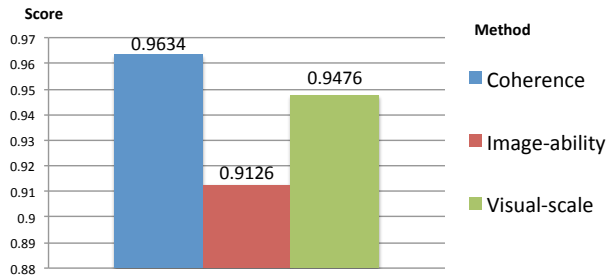


Figure 10: nDCG based evaluation for three criterion calculation methods.

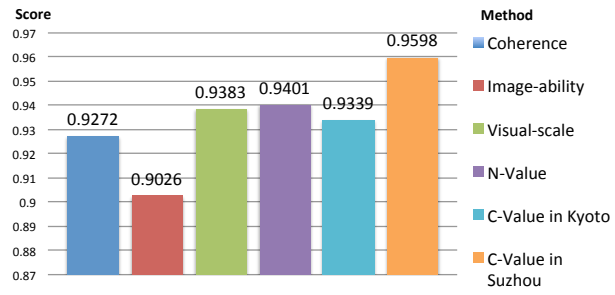


Figure 11: nDCG based evaluation for NV and CV estimation.

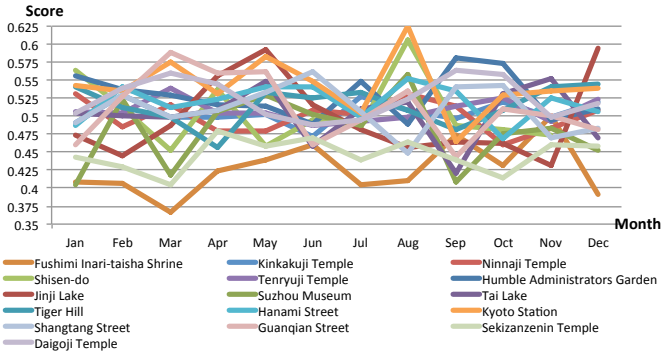


Figure 12: Monthly coherence scores based on color harmony.

perception. As a famous spot in Sekizanzenin Temple, the Sanjyusan Guanyin, which consists of 33 ordered and arranged avalokitevara (a type of Buddha), gives a high coherence score to this temple. This shows that our method can give a high score to a spot with harmonic color and repeated patterns, which satisfies the definition of coherence. The nDCG score for coherence calculation shown in Figure 10 is 0.9634.

The variation trend for color harmony and repeated patterns are shown in Figures 12 and 13. The lower the color harmony and repeated pattern scores, the higher the coherence held by the target spot.

According to the result for color harmony, the spots with many types of artificial architecture (i.e., Ninnaji Temple, Tenryuji Temple, and the Humble Administrators Garden) tend to maintain relatively stable scores throughout the whole year. The reason is that the changing of the seasons has little impact since tourists pay more attention to and take more photos of the artificial architecture than natural elements.

For the repeated patterns, the result shows that all the landscapes obtain smooth scores in a relatively fixed range except Jinji Lake. As mentioned previously, the clear sky and clear lake give Jinji Lake a high coherence for human perception. Besides, the regular light events held during certain festivals also fluctuate on the repeated pattern score because of their ability to attract peoples attention.

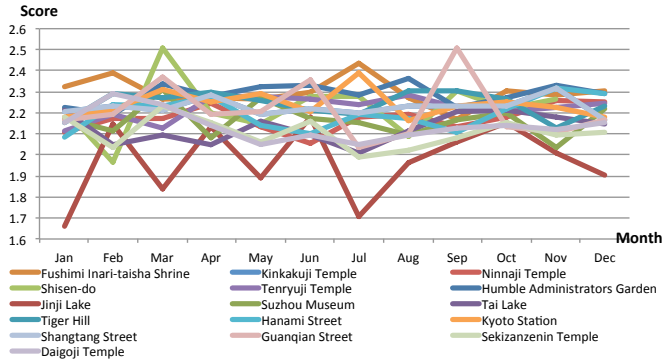


Figure 13: Monthly coherence scores based on repeated patterns.

The performance (nDCG) of nature evaluation implemented with only color harmony based on the time analysis is 0.9591 as shown in Figure 11.

4.2.2 Image-ability

By utilizing the method introduced in Section 3.1.2, the experimental results show in Figure 8 tell us that the method gives high scores to Shangtang Street, Jinji Lake, and Fushimi Inari Taisha Shrine, which matches their ground truths generated by the subjects. However, low scores are calculated for the Humble Administrators Garden even though its ground truth is quite high. One considerable reason is that the Humble Administrators Garden is famous for its classical Chinese architecture, so there are many non-landscape images included in the experimental data, such as images of interior decoration and interior design. Recall that the definition of image-ability is a strong visual image of a landscape that makes people have distinguishable and memorable experiences. Under this definition, outdoor aesthetic landscapes receive more attention than indoor ones. In contrast, a high score is calculated for Hanami Street even though its ground truth is quite low. This is explained by Hanami Street being famous for its night view, so the effect of bright lights may be treated as high image-ability in our method. The nDCG score for the image-ability calculation method is 0.9126.

Based on the experimental results shown in Figure 14, the monthly distributions for image-ability of each spot do not seem to have a regular pattern. Since the photo quality is affected by many factors (such as composition, objects, or even the focus of an image), it is difficult to determine whether a spot is beautiful or not just by considering photo quality. The nDCG result for image-ability shown in Figure 11 is 0.953.

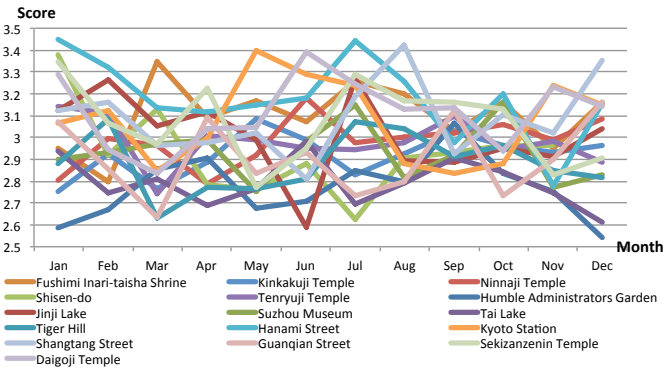


Figure 14: Monthly image-ability scores.

4.2.3 Visual-scale

Using the methods proposed by Oliva and Torralba (2001), we extract GIST features from each image and use CWM to estimate the openness and depth. Then we calculate a harmonic mean to determine the overall visual-scale score for a spot.

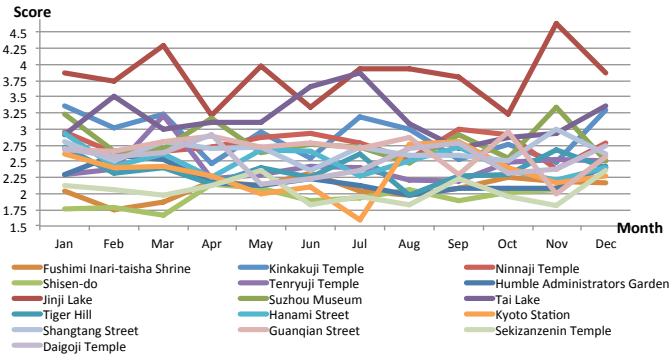


Figure 15: Monthly visual-scale scores.

According to the results shown in Figure 7, the calculated visual-scale scores for Jinji Lake and Tai Lake are clearly different from those of the other spots. The common feature for Jinji Lake and Tai Lake is that most images of them show wide lake scenery. Compared with the spots with a small space, this feature provides a stronger experience of wide-open appearance, which satisfies the definition of visual-scale in environmental psychology provided by Tveit et al. (2006). As shown in Figure 10, the nDCG score of visual-scale calculation is 0.9476.

The experimental results for visual-scale shown in Figure 15 indicate that all the spots maintain a smooth visual-scale score. The higher the visual-scale score, the higher the visual-scale held by the target spot. It is easy to explain that the visual-scale is a fixed criterion for a spot that does not vary over time. The nDCG score for visual-scale based nature evaluation shown in Figure 11 is 0.9627.

4.2.4 Evaluation on Spot Ranking

As an overall spot ranking based on NV estimation, our method combines three criteria (coherence, image-ability, and visual-scale) by calculating the normalized score for each criterion and taking the average score as the rank score for each spot. To the best of our knowledge, this is the first effort to rank sightseeing spots by utilizing environmental psychology criteria. The photo quality assessment we used in image-ability implementation is a basic method for sightseeing estimation even though its objective is different, so we consider this method as a baseline method for comparison.

We compare our method with two baseline methods. Since analyzing user rating data is one of the most common methods for spot ranking, we take the average score of users' ratings of TripAdvisor (http://www.tripadvisor.com/), as the rank score for each spot. For the second baseline, M. G. Berman et al. (2014) tested the relationship between low-level visual features with perceived naturalness and obtained results in which the non-straight edge density (NSED) has a strong correlation with perceived naturalness. Intuitively, we take the average value of NSED as the

rank score for each spot and compare it with our experiment result.

As shown in Figure 11, the nDCG score of combined NV is 0.9401, which is higher than that realized by using only image-ability. Figure 16 shows the nDCG score calculated with corresponding ground truth for User-Rating and NSED. The nDCG score for average score-based spot ranking in User-Rating based method 1 is 0.9341, and the NSED-based spot ranking in NSED based method is 0.8424.

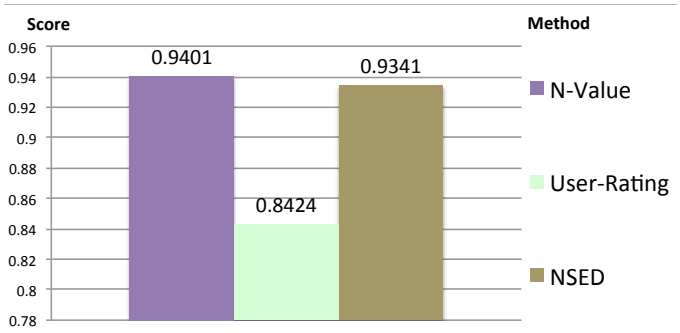


Figure 16: nDCG method for baseline: Nature Value

Our method achieves higher accuracy than the two basic methods. The User-Rating based method, which is realized by user-behavior, is not suitable for nature value estimation because the score given by users is affected by not only the quality but also the popularity of a spot. Based on the idea that naturalness tend to have an irregular shape, the NSED based method also gets an quite high accuracy in naturalness. Compared with the NSED method, our method consider more aesthetics factor for nature value estimation, which satisfies the needs for tourist.

4.2.5 Experimental Results and Discussion

In short, our method tends to assign a high score to spots with beautiful scenes, wide fields of vision, obvious color tendencies, or simple structures without being affected by popularity. However, as the Jinji Lake and Humble Administrators Garden case show, it seems that this rule is not appropriate for all the high nature spots perceived by people. Besides that, the content bias when taking a photo is another challenge for our method that should be solved.

The nDCG scores show that most sightseeing spots are ranked correctly. Specifically, we obtained the best performance when considering all three criteria. However, simply taking the average does not seem the best choice. In the future, we will investigate appropriate coefficients for each criterion.

According to the experiment results shown in Figures 7, 8, and 9, the criteria scores for Shisen-do, Sekizanzenin, and Daigoji Temple are calculated in a small error range. These results show that our method is effective for evaluating obscure spots, and an obscure spot with high sightseeing quality can be ranked higher than popular spots with low sightseeing quality.

In our experiment, our method gives a low NV to Kyoto Station, which matches the ground truth for nature perception. Intuitively, Kyoto Station has barely any NV or CV. Figure 17 shows representative photos of Kyoto Station. Most photos of the station are taken inside it and are filled with crowds. The experimental results indicate that our method can deal with this common case correctly and give a score lower than those for the other sightseeing spots.

Figure 17: Representative photos of Kyoto Station.

However, Jinji Lake obtains the highest combined NV even though its ground truth ranks behind those of Tiger Hill, Shisen-do, Tenryuji Temple, and Tai Lake. According to interviews with the subjects, one major reason for assigning a middle score to Jinji Lake is that although the major parts of the scene, i.e., the sky and lake, are nature elements, the buildings on the other side of the lake give a strong artificial perception to the whole spot. Figure 18 shows representative photos of Jinji Lake. However, our coherence based method highly evaluates this spot because of its simple structure and clear blue color tendency, and the methods for image-ability and visual-scale also give high scores for its beautiful lake view and broad field view, which leads to the high combined NV for Jinji Lake.

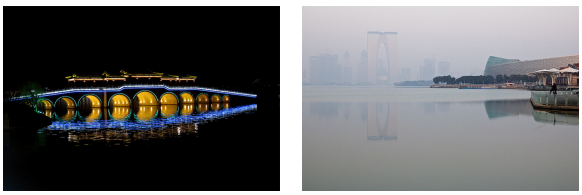


Figure 18: Representative photos of Jinji Lake.

The Humble Administrators Garden obtains a low NV even though it obtained a high ranked ground truth. As explained in Section 4.2.2, since tourists tend to pay more attention to interior decorations and interior designs than garden scenes, a large number of non-nature photos are taken, which lead to low scores for both the visual-scale and image-ability. Its coherence score is also low because of its complex indoor structure. Representative photos of the Humble Administrators Garden are shown in Figure 19.

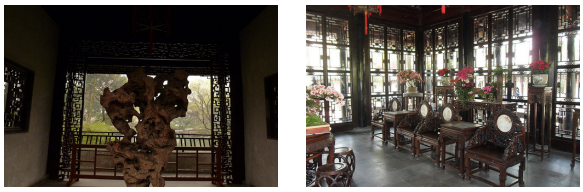


Figure 19: Representative photos of Humble Administrators Garden.

4.3 Evaluation on Culture Value

To evaluate the performance of the cultural value estimation, we first extract VGG features from images for each spot. Then, we use $k-$means to cluster images, which is simple but performs well. Images of each sightseeing spot are divided into 10 clusters. The threshold of step 2 in our method (see Section 3.2) is set to 100. We build an image graph for each cluster on the basis of this threshold and detect objects by looking for connected sub-graphs. In step 3, we download a training set from a search engine, which contains 450 images and 15 classes. This image set is used for training the DPM. The threshold of step 3 is set to 0. The nDCG result for the sightseeing spots from Kyoto and Suzhou are 0.9345 and 0.9603 as shown in Figure 11.

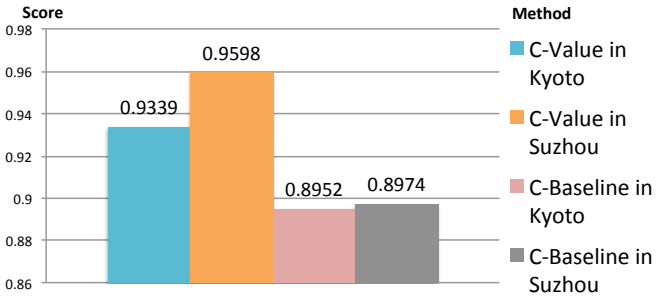


Figure 20: nDCG method for baseline: Culture Value

For the baseline, we go through all the sightseeing spots and calculate an average color for each city. The distance between the average colors of a sightseeing spot and a city can be taken as a metric for culture estimation. According to the method proposed by N. Hochman et al. (2012), the representative color tendency varies among cities of different cultures, which means that color tendency highly correlates with culture. Intuitively, in our baseline, the shorter the distance, the higher the culture value a sightseeing spot can be seen to hold. Figure 20 shows the nDCG scores for our C-Value method and baseline.

Our method gives Kinkakuji Temple a rank of 4 even though its true rank is 2. Kinkakuji Temple is more special than other spots. It is famous for a golden temple, which appears in many images taken there. Recall that our method detects objects in sightseeing spots. In this case, our method can only discover one object, which is in a large number of images taken by different tourists. Although we take the number of photos taken by different users into consideration, the lack of other objects still leads to a very bad rank for this spot. In our future work, we will make more effort to find better methods to solve these exceptions.

Besides, there are further challenges. For example, in addition to the scene images taken at a spot, there are also a lot of crowd images, food images, indoor images, and so on. Though some of them are filtered by Flickr's keyword based search, these noise images may affect our methods' performance.

4.4 Time complexity

The criteria algorithm for nature estimation goes through all the images only once, which means the cost increases linearly with the number of images. The culture estimation has the time complexity of $O(n^2)$ because a pairwise calculation for image similarity is needed in the second step. To demonstrate the applicability of our approach, we test our criteria algorithms separately by using four datasets with different sizes. The specifications of our experimental PC are: OS (Ubuntu 15.10), CPU (Intel i7 6770k, 4 cores), RAM (32 GB), and HDD (1.8 TB). We calculate each ranking criteria five times and show their average processing time in Table 3. Although we have not carried out an experiment on a large dataset, the results indicate that our methods can be applied to such a dataset because the quality estimation task is usually offline. However, since the algorithm is highly parallelizable, it can be appropriate for big data.

Table 3 Processing time for criterion calculation

No. of images	Coherence	Image ability	Visual scale	Culture Value
100	279.75s	1.85s	38.71s	291.84s
200	557.82s	3.40s	74.16s	697.99s
300	830.83s	4.67s	127.60s	1220.19s
400	1104.47s	6.17s	173.89s	1817.21s

5 Conclusions

In this paper, we presented novel methods to assess sightseeing value by analyzing geo-social images. We proposed three criteria for nature value (NV) assessment: coherence, image-ability and visual-scale. We also proposed a criterion for culture value (CV) assessment: architectural styles. Since the NV is affected by the time of year, we also developed a temporal analysis method for the NV. The experimental results demonstrated that our methods assess sightseeing value effectively.

For future work, we will try to improve our criterion calculation methods and find the relationship between criteria and sightseeing value.

Acknowledgment

This work is partly supported by JSPS KAKENHI Grant Numbers 25700033, 15J01402, and 16K12532.

References

Y. Zheng and X. Xie, "Learning travel recommendations from user-generated GPS traces," *ACM TIST*, vol. 2, no. 1, p. 2, 2011.

W.-C. Chen, A. Battestini, N. Gelfand, and V. Setlur, "Visual summaries of popular landmarks from community photo collections," in *ACM Multimedia*, 2009, pp. 789–792.

J. Liu, Z. Huang, L. Chen, H. T. Shen, and Z. Yan, "Discovering areas of interest with geo-tagged images and check-ins," in *ACM Multimedia*, 2012, pp. 589–598.

K. Hasegawa, Q. Ma, and M. Yoshikawa, "Trip tweets search by considering spatio-temporal continuity of user behavior," in *DEXA*, 2012, pp. 141–155.

Y.-T. Zheng, Z.-J. Zha, and T.-S. Chua, "Research and applications on georeferenced multimedia: a survey," *Multimedia Tools and Applications*, vol. 51, no. 1, pp. 77–98, 2011.

C. Zhuang, Q. Ma, X. Liang, and M. Yoshikawa, "Discovering obscure sightseeing spots by analysis of geo-tagged social images," in *ASONAM*, 2015, pp. 590–595.

C. Zhuang, Q. Ma, X. Liang, and M. Yoshikawa, "Anaba: An obscure sightseeing spots discovering system," in *ICME*, 2014, pp. 1–6.

S. Kaplan, R. Kaplan *et al.*, *Humanscape: Environments for people*. Duxberry press North Scitiate, MA, 1978.

S. C. Bourassa, *The Aesthetics of Landscape*. Behaven Press, London, 1991.

M. Tveit, Å. Ode, and G. Fry, "Key concepts in a framework for analysing visual landscape character," *Landscape Research*, vol. 31, no. 3, pp. 229–255, 2006.

J. Luo, D. Joshi, J. Yu, and A. Gallagher, "Geotagging in multimedia and computer visiona survey," *Multimedia Tools and Applications*, vol. 51, no. 1, pp. 187–211, 2011.

R. Ji, X. Xie, H. Yao, and W.-Y. Ma, "Mining city landmarks from blogs by graph modeling," in *ACM Multimedia*, 2009, pp. 105–114.

Y. Zheng, L. Zhang, X. Xie, and W.-Y. Ma, "Mining interesting locations and travel sequences from GPS trajectories," in *WWW*, 2009, pp. 791–800.

T. Hartig, "Nature experience in transactional perspective," *Landscape and Urban Planning*, vol. 25, no. 1, pp. 17–36, 1993.

A. E. Van den Berg, C. A. Vlek, and J. F. Coeterier, "Group differences in the aesthetic evaluation of nature development plans: a multilevel approach," *Journal of Environmental Psychology*, vol. 18, no. 2, pp. 141–157, 1998.

H. Ohta, "A phenomenological approach to natural landscape cognition," *Journal of Environmental Psychology*, vol. 21, no. 4, pp. 387–403, 2001.

X. Tang, W. Luo, and X. Wang, "Content-based photo quality assessment," *Multimedia, IEEE Transactions on*, vol. 15, no. 8, pp. 1930–1943, 2013.

R. Datta, D. Joshi, J. Li, and J. Z. Wang, "Studying aesthetics in photographic images using a computational approach," in *ECCV*, 2006, pp. 288–301.

S. Dhar, V. Ordonez, and T. L. Berg, "High level describable attributes for predicting aesthetics and interestingness," in *CVPR*, 2011, pp. 1657–1664.

W. Yin, T. Mei, and C. W. Chen, "Assessing photo quality with geo-context and crowdsourced photos," in *VCIP*, 2012, pp. 1–6.

M. G. Berman, M. C. Hout, O. Kardan, M. R. Hunter, G. Yourganov, J. M. Henderson, T. Hanayik, H. Karimi, and J. Jonides, "The perception of naturalness correlates with low-level visual features of environmental scenes," *PloS one*, vol. 9, no. 12, 2014.

M. R. Hunter and A. Askarinejad, "Designer's approach for scene selection in tests of preference and restoration along a continuum of natural to manmade environments," *Frontiers in psychology*, vol. 6, p. 1228, 2015.

B. De Mente, *Elements of Japanese Design*. Tuttle Publishing, 2011.

P. Emmons, J. Lomholt, and J. Hendrix, *The cultural role of architecture: contemporary and historical perspectives*. Routledge, 2012.

R. Harrold and P. Legg, *Folk Costumes of the World*. Cassell Illustrated, 1999.

S. Pendergast, T. Pendergast, and S. Hermsen, *Fashion, Costume, and Culture*. UXL, 2003.

A. C. Berg, F. Grabler, and J. Malik, "Parsing images of architectural scenes," in *ICCV*, 2007, pp. 1–8.

A. Toshev, P. Mordohai, and B. Taskar, "Detecting and parsing architecture at city scale from range data," in *CVPR*, 2010, pp. 398–405.

Z. Xu, D. Tao, Y. Zhang, J. Wu, and A. C. Tsoi, "Architectural style classification using multinomial latent logistic regression," in *ECCV*, 2014, pp. 600–615.

L. Bossard, M. Dantone, C. Leistner, C. Wengert, T. Quack, and L. Van Gool, "Apparel classification with style," in *ACCV*, 2012, pp. 321–335.

Y. Matsuda, "Color design," *Asakura Shoten*, 1995.

T. Kohonen and P. Somervuo, "Self-organizing maps of symbol strings," *Neurocomputing*, vol. 21, no. 1, pp. 19–30, 1998.

J. Harel, C. Koch, and P. Perona, "Graph-based visual saliency," in *NIPS*, 2006, pp. 545–552.

S. Guiaşu, "Weighted entropy," *Reports on Mathematical Physics*, vol. 2, no. 3, pp. 165–179, 1971.

N. Murray, L. Marchesotti, and F. Perronnin, "AVA: A large-scale database for aesthetic visual analysis," in *CVPR*, 2012, pp. 2408–2415.

N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in *CVPR*, 2005, pp. 886–893.

M. A. Stricker and M. Orengo, "Similarity of color images," *IProc. SPIE Storage and Retrieval for Image and Video Databases*, vol.2420, 1995, pp. 381–392.

T. Ojala, M. Pietikäinen, and D. Harwood, "A comparative study of texture measures with classification based on featured distributions," *Pattern recognition*, vol. 29, no. 1, pp. 51–59, 1996.

A. Oliva and A. Torralba, "Modeling the shape of the scene: A holistic representation of the spatial envelope," *International Journal of Computer Vision*, vol. 42, no. 3, pp. 145–175, 2001.

N. Hochman and R. Schwartz, "Visualizing Instagram: Tracing cultural visual rhythms," in *SocMedVis*, 2012, pp. 6–9.

K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," in *arXiv*, 2014, pp. 1–14.

D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *International Journal of Computer Vision*, vol. 60, no. 2, pp. 91–110, 2004.

P. F. Felzenszwalb, R. B. Girshick, D. McAllester, and D. Ramanan, "Object detection with discriminatively trained part-based models," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 32, no. 9, pp. 1627–1645, 2010.

K. Jarvelin and J. Kekalainen, "Cumulated gain-based evaluation of IR techniques," ACM Transactions on Information Systems, vol. 20, pp. 422–446, 2002.

M. G. Berman, M. C. Hout, and O. Kardan, "The perception of naturalness correlates with low-level visual features of environmental scenes," PloS one, pp. 9–12, 2014.

N. Hochman and R. Schwartz "Visualizing Instagram: Tracing cultural visual rhythms," International AAAI Conference on Weblogs and Social Media, pp. 6–9, 2012.