JOURNAL OF INFORMATION SCIENCE AND ENGINEERING 30, 619-635 (2014)

DOI:10.1688/JISE.2014.30.3.5

Hybrid Approach of Situation-Oriented Classification of Sightseeing Spot Images Based on Visual and Tag Information

CHIA-HUANG CHEN AND YASUFUMI TAKAMA
Graduate School of System Design
Tokyo Metropolitan University
Tokyo, 192-0397 Japan
E-mail: chen-chiahuang@ed.tmu.ac.jp; ytakama@tmu.ac.jp

Recent trend on the web is to share their traveling experience via uploading photos to web albums. Shared photos of sightseeing spots are important resources for those who are going to visit there. As sightseeing spot scenes vary with different situations, such as weather and season, automatic classification of photos into different situations is expected to be beneficial for tourists to plan when to visit there. This paper proposes a hybrid approach of combining content-based image classification with filtering based on tag information of image. By using geotag information when retrieving images from web albums, collected images can be limited to a reasonable boundary to eliminate outliers. Content-based image classification groups collected images into night, sunrise/sunset, cloudy, and shine situations. Moreover, by using the timestamp of images, the four situation categories are further verified to increase the accuracy. Experimental results show that the hybrid approach of content-based image classification and tag-based filtering is effective for classifying image into situations with high precision and recall.

Keywords: classification, color feature extraction, geotag, timestamp, tourism informatics

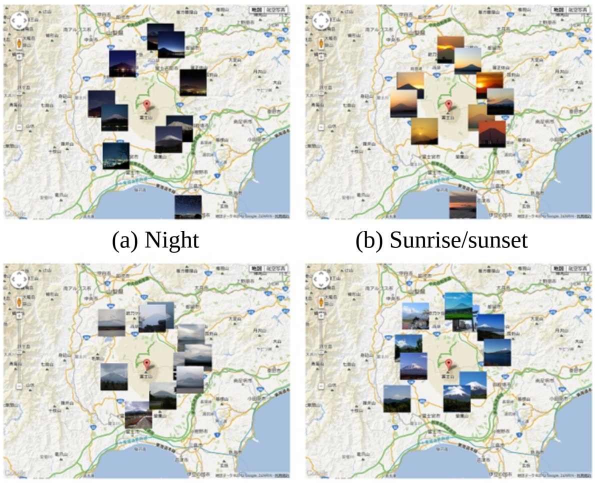
1. INTRODUCTION

This paper proposes a hybrid method which combines content-based classification and filtering with tag information for grouping sightseeing spot images into categories of different situations. Recently, web albums such as Flickr and Picasa become popular, to which tourists upload taken photos for online sharing. These images are useful for other people who are interested in visiting there. Generally speaking, the impression of a sightseeing spot highly depends on a situation such as weather condition and season. For example, some spots are famous for its night view. Natural sceneries such as mountains, rivers, and gardens vary with seasons. Therefore, providing users with images of sightseeing spots with respect to each situation will help them planning their trips.

Based on this idea, we are developing Web-based tourist service that shows how scenery changes depending on the situation. Fig. 1 shows a screenshot of the prototype tourist service currently under development, in which only images of a certain situation are mapped based on geotag information. The system is expected to be useful to users deciding when to visit which sightseeing spots.

Such a tourist service requires a method for classifying various photos available on the Web into situations. There have been many effective methods for image retrieval, some of which are applied to web image search [2, 5, 7, 10, 17]. However, main focus of

Received February 28, 2013; accepted June 15, 2013.
Communicated by Hung-Yu Kao, Tzung-Pei Hong, Takahira Yamaguchi, Yau-Hwang Kuo, and Vincent Shin-Mu Tseng.



(a) Night (b) Sunrise/sunset

(c) Cloudy (d) Shine situations

Fig. 1. Prototype of web-based tourist service system.

image retrieval has been on finding similar images for general purposes. In order to calculate similarity between images in terms of situations, more specific approach considering characteristics of situations is required.

Because of the spread of digital camera and GPS devices, the photos shared on web albums contain not only image files but also those tag information such as EXIF (exchangeable image file format), geotag, and timestamp. This kind of information is useful for location and time filtering. Therefore, the proposed hybrid method combines a content-based image classification method and image filtering based on tag information. The image classification method employs hierarchical approach [3]. In each stage color features extracted from ROI (region of interest) are used for identifying clusters of a certain situation. As it is supported that each of the target four situations corresponds to a specific time of day, image's timestamp becomes an important clue for verification of clusters obtained by the content-based image classification. Using timestamp information can filter out the image which was mis-classified in unsuitable situation such as a shine image mis-classified in night situation. Furthermore, in the phase of image collection, the unsuitable images such as those taken far away from or inside the target spot can be filtered out by geotag information for constructing a reliable image dataset.

Experiments are conducted to evaluate the performance of the proposed method. Test dataset consists of photos which are collected from Flickr and labeled as night, sunrise/sunset, cloudy, and shine situations manually. It contains total 1294, 965, 398, 553, 304, 399, and 499 images of Tokyo Tower, Mt. Fuji, Daiba, Sensoji, Meiji Shrine, Rainbow Bridge Tokyo, and Arashiyama respectively. The experimental results show that for various kinds of sightseeing spots, proposed hybrid approach improve performance.

This paper is an extended version of a paper [4]. Main difference is that 5 other spots are newly evaluated in the experiments, and sun rising and setting times of Kyoto is added for new spots. The performance of image processing part is also improved by modifying color components used in part of the algorithm. Furthermore, the advantages of the proposed method are discussed in more detail through comparison with content-based method and timestamp only method.

The rest of this paper is organized as follows. Section 2 briefly reviews the existing works related to feature extraction, image classification method, and utilization of tag information for image clustering. Section 3 describes the proposed method. In section 4, experimental results are presented to evaluate the effectiveness of our approach.

2. RELATED WORKS

Since the amount of available images grows up rapidly, the effective grouping of vast amount of images into meaningful classes can be useful for many kinds of applications such as indexing of image database [14], categorization of traveling images [15], and browsing of video shots [19]. Vailaya *et al.* [14] have attempted to use binary Bayesian classifiers and capture high-level concepts from low-level image features for the hierarchical classification of vacation images. The images are classified into indoor and outdoor classes at the first level, and outdoor images are further grouped into city and landscape classes. Finally the landscape images are classified into sunset, forest, and mountain classes.

Different to supervise learning method [16], clustering method can group a set of unsupervised (unlabeled) data into several clusters based on their low-level visual features. Silakari *et al.* [13] have focused on color feature of images. The color moment and Block Truncation Coding (BTC) are used to extract features and *K*-means clustering algorithm is applied to group 1,000 images into 10 clusters such as busses, dinosaurs, and flowers. Sleit *et al.* [12] have utilized the color histogram, Gabor filters, and Fourier transformation for color, texture, and shape feature extraction, respectively to group images based on K-means clustering. The resultant image database includes four different groups which consist of dinosaurs, flowers, busses and elephants. Huang [6] has integrated the local SIFT (Scale Invariant Feature Transformation) feature with the global CLD (Color Layout Descriptor) feature and adopted the affinity propagation clustering algorithm which does not need to initialize the number of clusters. Furthermore, the bag of visual word model is applied in re-clustering for the enhancement of clustering performance. The dataset consists of 750 groups of 4 images each.

In order to establish a clustering technique that can handle the massive amounts of user-generated photos, Papadopoulos *et al.* [11] have proposed image similarity graph that is constructed based on both visual and tag features and applied the community detection to efficiently identify clusters of images. Moëllic *et al.* [8] have proposed a clustering approach based on the shared nearest neighbors algorithm (SNN), which employs both textual data (tags) and visual features to build representative clusters. Its evaluation is conducted with 1,000 images classified in ten well separated categories.

3. PROPOSED METHOD

On the visual perspective, it is supposed that people consider the images belong to a certain situation according to the changes of brightness or color in local region. Therefore, local color feature extraction is expected to be meaningful for classifying images into situations. For example, shine and sunset images have blue and orange colors in sky area respectively. At the same time, different situations such as night, sunrise/sunset, and daytime (cloudy and shine), images corresponds to different time of day. For example, night images are taken after evening and before morning, while sunrise/sunset images are taken at either morning or evening. Based on this consideration, this paper proposes a hybrid approach for clustering images into situations. The proposed method consists of several stages, in each of which after applying content-based image classification, tag-based filtering is applied to improve the accuracy of clustering.

3.1 Overall Procedure

Fig. 2 shows a hierarchical organization applied in this paper for the purpose of grouping images into categories of different situations. At first sightseeing spot images are collected from web album such as Flickr. In this phase, most of images are supposed to be distinguished into indoor scenes, outdoor scenes, and others such as faces with using existing method [14]. Geotag information is also used to filter out photos taken inside or far away from the target spot. After applying such methods, only outdoor photos of target sightseeing spots are used as input images of the proposed method.

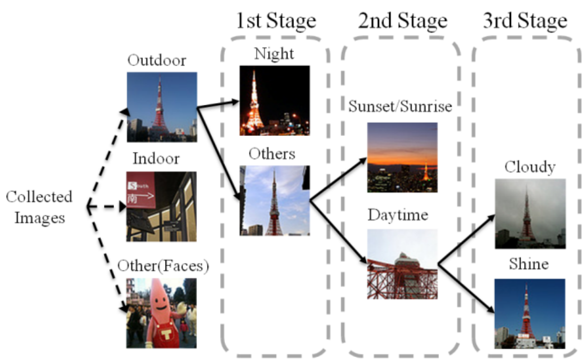


Fig. 2. Hierarchical organization of situation categories.

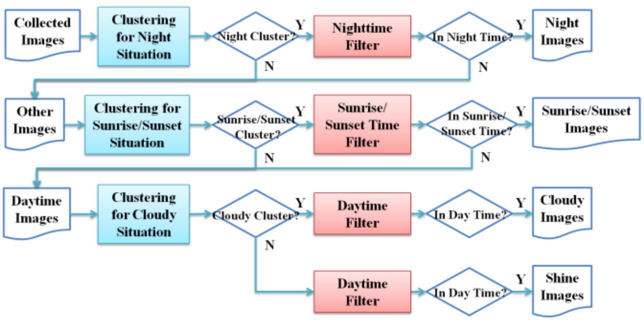


Fig. 3. Overall processing flow.

The overall processing flow as shown in Fig. 3 consists of three clustering and three filtering processes. In the collected image dataset, the classification of different situations by time information only is too rough because the sunrise and sunset time varies with the season. Moreover, there are some images with wrong shooting time. If time filtering process is applied prior to clustering, these kinds of image will be ignored and cause the result of clusters corresponding to specific situation to get worse. Therefore the content-based image classification method [3] is applied to divide the collected images into night cluster and other images at the beginning. By using tag information, the images in night cluster are verified with the nighttime filter, which separates mis-clustered images from night images. Other images which are not classified in night cluster and the mis-clustered images will be considered as input for next clustering process. This round of process corresponds to the 1st stage as shown in Fig. 3. The other images are further divided into sunrise/sunset and daytime clusters at 2nd stage by applying content-based image classification, which is followed by time filter for sunrise/sunset is applied. At the 3rd stage, the cloudy images are discriminated from shine images by content-based classification and daytime filter for verification.

Table 1. Sun rising and setting times in the first day of each month at Tokyo, Shizuoka, and Kyoto.

Year/Month	Tokyo		Shizuoka		Kyoto	
	Rising	Setting	Rising	Setting	Rising	Setting
2011/01	6:50	16:38	6:54	16:45	7:05	16:56
2011/02	6:42	17:80	6:46	17:15	6:56	17:25
2011/03	6:12	17:36	6:17	17:41	6:27	17:52
2011/04	5:29	18:02	5:34	18:07	5:45	18:18
2011/05	4:50	18:27	4:57	18:31	5:07	18:42
2011/06	4:27	18:51	4:34	18:55	4:44	19:05
2011/07	4:28	19:01	4:36	19:04	4:46	19:15
2011/08	4:48	18:46	4:55	18:50	5:06	19:01
2011/09	5:12	18:10	5:18	18:14	5:29	18:25
2011/10	5:35	17:26	5:40	17:32	5:51	17:42
2011/11	6:02	16:47	6:07	16:53	6:17	17:04
2011/12	6:31	16:28	6:35	16:35	6:46	16:46

3.2 Definition of Time Window

The four situations handled in this paper are night, sunrise/sunset, cloudy, and shine as mentioned above. As different situations correspond to different time of day, except that cloudy and shine situation, this paper sets time windows for each situation in order to filter out unsuitable images from clusters obtained by content-based image classification. As such time windows are supposed to change according to seasons because of the change of sun rising and setting times, we investigated those times of target sightseeing spots (Tokyo, Shizuoka, and Kyoto in this paper) from the website of National Astronomical Observatory of Japan[1]. Table 1 shows sun rising and setting times in the first day of each month. In case of Tokyo Tower, Daiba, Sensoji, Meiji Shrine, and Rainbow Bridge Tokyo, sun rising and setting times of Tokyo are applied. Sun rising and setting times of Shizuoka are applied for Mt. Fuji. In the case of Arashiyama, it refers to the times of Kyoto.

[1] http://eco.mtk.nao.ac.jp/cgi-bin/koyomi/koyomix_en.cgi

Table 2. Time window as filters for different situations in April.

Situation	Time Window		
	Tokyo	Shizuoka	Kyoto
Sunrise	3:00 ~ 7:00	4:00 ~ 8:00	4:00 ~ 8:00
Daytime (Cloudy & Shine)	3:00 ~ 20:00	4:00 ~ 20:00	4:00 ~ 20:00
Sunset	16:00 ~ 20:00	16:00 ~ 20:00	16:00 ~ 20:00
Night	1:00 ~ 7:00, 16:00 ~ 24:00	1:00 ~ 8:00, 16:00 ~ 24:00	1:00 ~ 8:00, 16:00 ~ 24:00

Considering such seasonal variation and the influence of weather conditions, the proposed method employs different overlapping time windows in each month. For example, time windows for April are shown in Table 2. The range of each time window in sunrise and sunset is 4 hours. It is noted that time windows for daytime and night time are set to include time windows of sunrise and sunset.

3.3 Processing at 1st Stage

The goal of this stage is to discriminate night images from other images. In human perception, the darkness and brightness are commonly used for recognition of daytime and night [18]. Thus the brightness is useful for discriminating night images from others. At this stage, the histogram of value component within the top one-third region of an image is calculated as local color feature for K-means clustering. The situation discrimination is applied to separate the night and other clusters. The detailed description of the processing is given in [3].

After clustering and discrimination process, the nighttime filter considering the time window of night situation is applied to verify night cluster's images. As noted in Section 3.2, time window to be applied is selected according to the shooting date of an image. This process is expected to filter out images which contain very deep blue sky area but taken in daytime, or those of which top region are covered by some object like plants. Such outliers will be merged with non-night cluster as input for next stage.

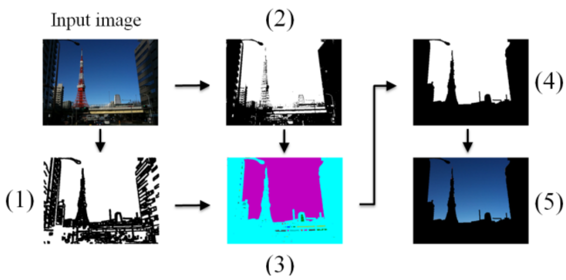


Fig. 4. Segmentation flow of ROI.

3.4 Processing at 2nd Stage

In this stage, a further grouping of images, which are contained in another cluster than night one, into sunrise/sunset and other situations is considered. In order to extract sky region containing no other objects as ROI, a region segmentation method using edge detection is employed [3]. The method consists of the following 5 steps, which are also shown in Fig. 4.

(1) Apply Canny edge detection [1] to obtain edge region (R_e) of an input image and then reverse the dilated R_e which is dilated with 5×5 kernel by morphology operation to obtain non-edge region (R_{ne}).
(2) Get a global image threshold of value component by Otsu's method [9] and convert input image to binary image.
(3) Get a binary image as the intersection of R_{ne} and the binary image from step (2) and then apply 8-connectivity on it to obtain the connected regions.
(4) Get the binary image of maximal region and then obtain ROI by dilating the binary image with 5×5 kernel.
(5) Extract the color feature within ROI.

After applying these series of steps, the color histogram is calculated by extracting the Cb and Cr components of YCbCr within the ROI for clustering. Based on preliminary experimental results, the color features are changed from the combination of RGB and CbCr components to CbCr only in order to improve the performance from the method as proposed in [4].

Figs. 5 and 6 show the examples of obtained 8 clusters including sunrise/sunset ones. The figure consists of sample images in the clusters and histograms of their centroids on bins of Cb and Cr. It is seen that sunrise/sunset cluster has smaller peak than other clusters. Therefore, a cluster having the smallest peak in the histogram of a centroid is selected as sunrise/sunset cluster.

After clustering and discrimination process, the time filter considering the time window of sunrise/sunset situation such as shown in Table 2 is applied to verify the taken time of images in sunrise/sunset cluster to improve the accuracy in this stage.

3.5 Processing at 3rd Stage

The purpose of the final stage is to group remaining images from the 2nd stage into cloudy and shine situation. The ROI is segmented by the same method as the 2nd stage. The Cb and Cr components are employed as features also in this stage for clustering. After *K*-means clustering, peak values between 16 and 32 of Cb component of clusters' centroids are compared with mean value of peak values among all clusters. The cluster of which centroid has the higher peak values in Cb component than the mean value is selected as a cluster of cloudy situation. When multiple clusters satisfy the condition, those are merged into one cluster. The rest of the extracted clusters are also merged and considered as shine cluster. It is noted that the discrimination criterion is changed from Cr component to Cb component based on preliminary experimental results, in order to improve the performance from the method as proposed in [4].

After clustering and discrimination process, the time filter considering the time window of daytime situation such as shown in Table 2 is applied to verify the taken time of images in both cloudy and shine clusters to improve the accuracy in this stage.

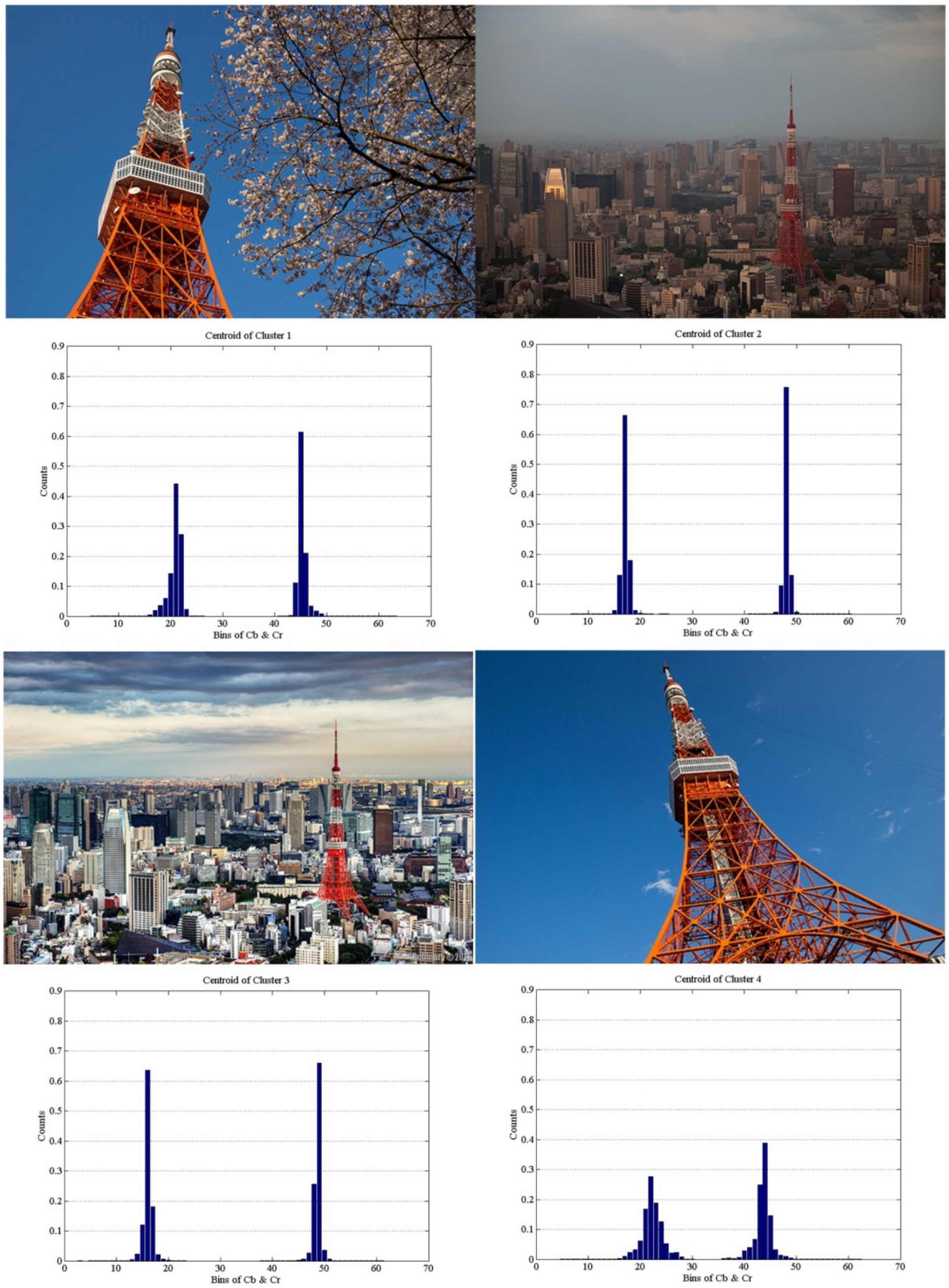


Fig. 5. Sample images in obtained cluster 1 to 4 and histogram of Cb & Cr values of their centroids at 2nd stage.

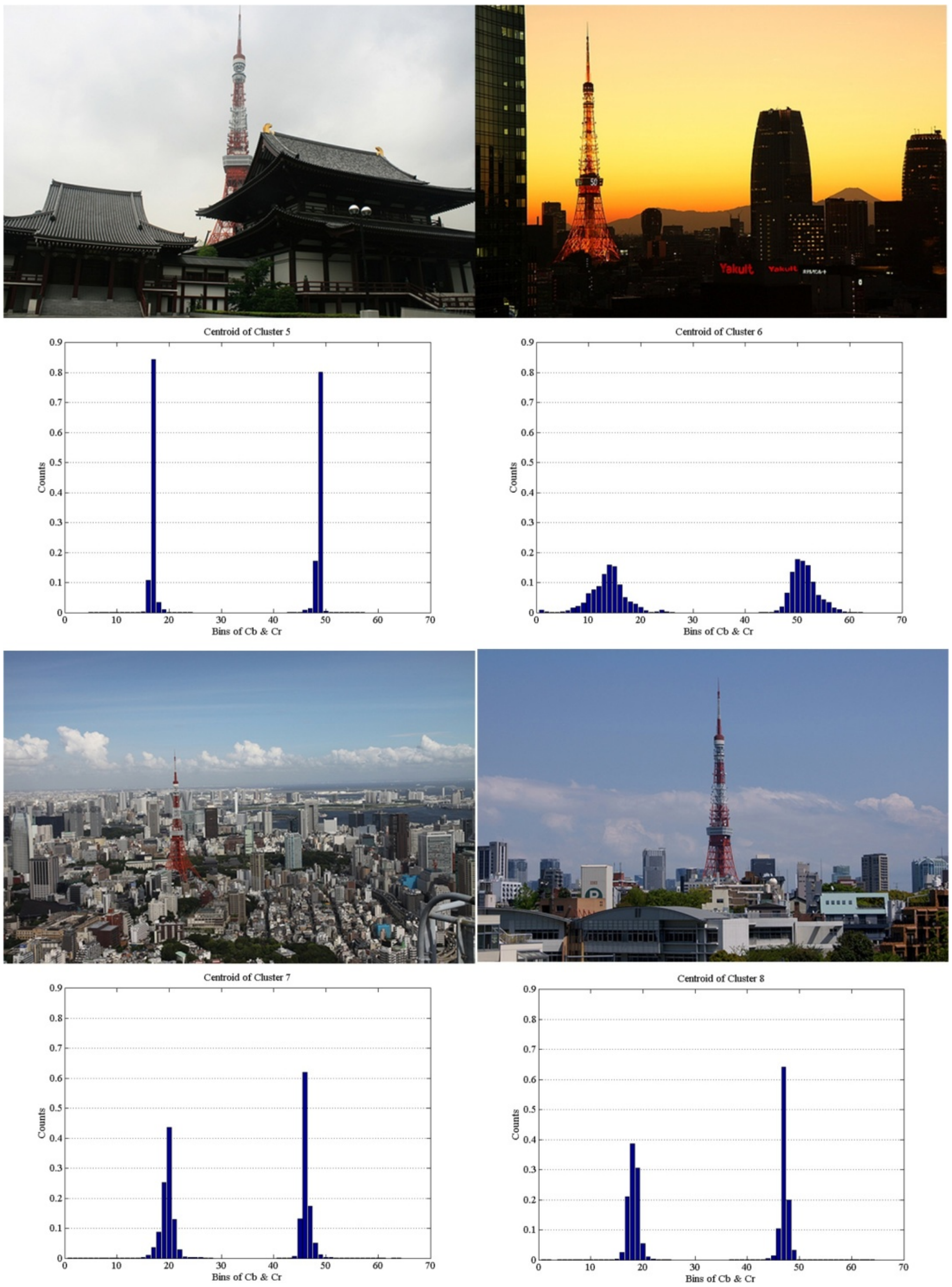


Fig. 6. Sample images in obtained cluster 5 to 8 and histogram of Cb & Cr values of their centroids at 2nd stage.

Table 3. Parameter settings of image collection.

Query Text	Central Point (Latitude, Longitude)	Collection Range	Filter Range	Total Images
Tokyo Tower	35.658610, 139.745447	0.2	0.00005	3,792
Mt. Fuji	35.362596, 138.731232	0.4	0.07	2,409
Daiba	35.630599, 139.778449	0.05	0.0003	1,297
Sensoji	35.714751, 139.796685	0.05	0.0002	2,050
Meiji Shrine	35.676324, 139.69938	0.05	0.0003	1,981
Rainbow Bridge Tokyo	35.635604, 139.76635	0.2	0.00005	608
Arashiyama	35.015194, 135.677706	0.2	0	3,700

4. EXPERIMENTS

Experiments are conducted in order to evaluate the performance of the proposed method. Comparison of hybrid method, content-based classification, and using timestamp only is conducted to evaluate the effectiveness of proposed hybrid method.

The proposed method is implemented on Matlab and Java. We collected images of 7 sightseeing spots with Flickr by setting a bounding box and recorded their geotag and shooting timestamp together with images. By using the geotag, the limited boundary was constructed to filter out unsuitable images which were taken inside or far away from the target spots. The parameter settings such as query text, geo degrees of central point, collection range, and filter range for image collection are shown in Table 3, together with total number of collected images after such filtering. For example, the image set of Tokyo Tower was collected according to boundary of latitude (35.658610 ± 0.2) and longitude (139.745447 ± 0.2) and then filter boundary of latitude (35.658610 ± 0.00005) and longitude (139.745447 ± 0.00005) was applied to eliminate the images which were taken inside the tower. The filter range of each sightseeing spot is determined based on its geographical size. That is, small value is used for specific construction such as Tokyo Tower and Rainbow Bridge Tokyo. On the other hand, the largest value is used for Mt. Fuji.

Table 4. Test dataset.

Spot	Night	Sunrise/ Sunset	Cloudy	Shine	Total Labeled Images
Tokyo Tower	577	64	248	405	1,294
Mt. Fuji	40	92	236	597	965
Daiba	118	69	57	154	398
Sensoji	93	17	226	217	553
Meiji Shrine	42	4	145	113	304
Rainbow Bridge Tokyo	149	94	50	106	399
Arashiyama	54	39	226	180	499

Various kinds of sightseeing spots with different characteristics such as natural scene and artificial scene are selected for evaluating the effectiveness and limitation of proposed method. The images correspond to a situation (night, sunrise/sunset, cloudy and shine) were selected and labeled manually. Two people took part in the labeling process, and an image is labeled only when both of them agree to the label. Table 4 summarizes the obtained test dataset.

In order to evaluate the performance of the proposed method, we apply the measures of precision and recall commonly used in information retrieval. The precision (Eq. (1)) is measured by computing the ratio of number of relevant images in a cluster divided by the total number of images in the cluster. The recall (Eq. (2)) is computed by dividing the number of relevant images in a cluster by the total number of relevant images in the dataset. The *F*-measure (Eq. (3)) is a balanced mean between precision and recall.

$$precision = \frac{\#\,of\ \ relevant\ \ images\ \ in\ \ a\ \ cluster}{\#\,of\ \ images\ \ in\ \ a\ \ cluster} \tag{1}$$

$$recall = \frac{\#\,of\ \ relevant\ \ images\ \ in\ \ a\ \ cluster}{\#\,of\ \ relevant\ \ images} \tag{2}$$

$$F = 2 \times \frac{precision \times recall}{precision + recall} \tag{3}$$

In order to explain the characteristics of employed content-based image classification, the result of the first stage in Fig. 3 is shown. Table 5 shows the experimental result in night situation at the 1st stage for Tokyo Tower and Mt. Fuji. Different color feature extraction is applied to construct the night cluster. The values of precision and recall are measured for following methods.

Table 5. Precision and recall values (%) of night situation at 1st stage.

Spot	Measures	Local Value	Local Intensity	Global Value	Global Intensity
Tokyo Tower	Precision	98.99	98.37	95.70	97.12
	Recall	84.58	83.54	84.92	81.80
Mt. Fuji	Precision	82.50	43.90	12.59	9.97
	Recall	82.50	90.00	87.50	92.50

- Local value: value component of HSV space within top 1/3 region of image. (employed in proposed method)
- Local intensity: intensity within top 1/3 region of image.
- Global value: value component of HSV space within whole image.
- Global intensity: intensity within whole image.

It is seen that the value component performs better than intensity. It is observed that the precision of intensity gets worse when a shine image contains deep blue or dark sky. Fig. 7 shows an image of Tokyo Tower in shine and night situations. Intensity of both images is similar, whereas value is distinguishable.

It is also seen the result of local color feature is better than global color feature. Although there is no big difference in recall between local and global features, precision of local feature is much higher than global feature. As the number of images of a sight-

Fig. 7. The comparison of intensity and value component.

seeing spot obtained from the Web is usually huge, we consider the precision is more important than recall.

In order to evaluate the proposed hybrid method, the values of precision and recall are measured and compared with other two methods.

- Timestamp Only: this method verifies four situations by using only time windows (*i.e.* without content-based image classification). Different time window is applied to each image according to its shooting date.
- Content-based: this method skips filtering processes with using time windows in Fig. 3.
- Hybrid (proposed method): this method performs content-based image classification first and then utilizes time windows to filter out outliers as shown in Fig. 3.

Proposed method uses *K*-means clustering in each stage. That means result of execution is different in each time. Therefore, the experiment performs *K*-means 10 times for each stage and then calculates average precision and recall. Table 6 compares average precision and recall of hybrid approach (proposed method) and other two methods mentioned above. Precision values when using timestamp only is calculated as the ratio of correctly labeled images among all images within the corresponding time window. In most cases, the proposed method can get the best results in precision. On the other hand, it is seen that recall of hybrid method tends to be lower than content-based method. It is because time filter filters out not only irrelevant but also relevant images. However, better result in precision shows that more irrelevant (mis-clustered) images are filtered out as outliers. One of typical cases where the time filter is effective is sunrise/sunset situation in Arashiyama, of which precision improves about 37 points with only 8 point decrease of recall.

Comparison between the content-based and timestamp only shows content-based approach is effective. Using timestamp only suffers from the worse precision in all of four situations. Although the time of sun rising and setting can be defined by the altitude of sun, the actual daytime and nighttime vary with position of a spot and season. Therefore, it is observed that many night and shine images are contained in sunrise/sunset time window. From these results, it can be said that using timestamp information as a means for supplementing the performance of content-based image classification, which is our proposed approach, is reasonable.

Table 6. Average precision and recall values (%) measured by different methods in each situation.

Spot	Method	Night (Precision /Recall)	Sunrise/Sunset (Precision /Recall)	Cloudy (Precision /Recall)	Shine (Precision /Recall)
Tokyo Tower	Timestamp Only	59.47/97.92	7.55/85.94	22.85/98.79	37.22/98.52
	Content-based	98.08/88.73	48.28/87.05	72.54/77.50	78.15/84.79
	Hybrid	99.01/87.00	61.38/75.00	73.69/76.29	80.12/83.55
Mt. Fuji	Timestamp Only	6.77/100	14.56/89.13	25.19/100	63.18/99.16
	Content-based	65.38/85.00	71.05/58.70	74.20/78.90	88.37/89.70
	Hybrid	79.07/85.00	78.46/55.43	74.20/78.90	88.72/88.86
Daiba	Timestamp Only	47.97/100	27.14/82.61	15.19/96.49	41.71/98.05
	Content-based	94.83/93.22	89.59/67.97	86.36/66.67	75.13/96.10
	Hybrid	95.65/93.22	92.57/54.20	85.71/63.16	76.72/94.16
Sensoji	Timestamp Only	31.05/82.80	5.02/64.71	42.75/99.12	41.03/99.08
	Content-based	88.30/89.25	10.24/58.82	92.09/90.00	83.54/85.72
	Hybrid	92.00/74.19	15.33/35.29	93.36/90.00	84.23/85.26
Meiji Shrine	Timestamp Only	30.60/97.62	2.65/75.00	49.12/95.86	37.10/92.92
	Content-based	75.47/95.24	6.45/50.00	82.88/93.51	89.25/68.85
	Hybrid	86.67/92.86	12.50/25.00	89.24/90.76	91.26/67.97
Rainbow Bridge Tokyo	Timestamp Only	52.46/100	37.50/89.36	14.45/98.00	30.97/99.06
	Content-based	95.27/94.63	99.28/69.15	70.02/86.80	77.59/90.38
	Hybrid	95.92/94.63	100/67.02	72.09/84.80	78.54/89.43
Arashiyama	Timestamp Only	26.21/100	20.81/92.31	46.57/96.02	36.05/93.33
	Content-based	90.00/100	47.62/76.92	90.67/88.36	82.99/87.44
	Hybrid	94.74/100	84.38/69.23	91.69/84.69	82.95/82.55

Table 7. The comparison of hybrid method and content-based method in *F*-measure (%).

Spot	Method	Night (*F*-measure)	Sunrise /Sunset (*F*-measure)	Cloudy (*F*-measure)	Shine (*F*-measure)
Tokyo Tower	Content-based	73.91	64.29	76.48*	89.03*
	Hybrid	81.93*	64.96*	76.48*	88.79
Mt. Fuji	Content-based	93.17*	62.23	74.94	81.33
	Hybrid	92.62	67.51*	74.97*	81.80*
Daiba	Content-based	94.02	77.30*	75.25*	84.33
	Hybrid	94.42*	68.37	72.73	84.55*
Sensoji	Content-based	88.77*	17.44	91.03	84.62
	Hybrid	82.14	21.37*	91.65*	84.74*
Meiji Shrine	Content-based	84.21	11.43	87.87	77.73
	Hybrid	89.66*	16.67*	89.99*	77.91*
Rainbow Bridge Tokyo	Content-based	94.95	81.52*	77.51	83.5
	Hybrid	95.27*	80.25	77.93*	83.63*
Arashiyama	Content-based	94.74	58.82	89.50*	85.16*
	Hybrid	97.30*	76.06*	88.5	82.75

In cloudy situation of Daiba and shine situation of Arashiyama, average precision and recall of hybrid method are slightly lower than content-based method. That is because the number of cloudy images with wrong taken time is more than the number of irrelevant images. Therefore, time windows filtered out too many relevant images and cause worse result.

In order to compare overall performance between content-based and hybrid methods, Table 7 shows *F*-measure. In the table, the better result is marked with asterisk (*). It is seen the proposed method can get better result than the content-based method for all 4 situations in Meiji Shrine. The proposed method can also get better result than the content-based method for 3 situations in Tokyo Tower, Mt. Fuji, Sensoji, and Rainbow Bridge Tokyo.

Comparison of the results among different situations shows that much worse results than other situations are sometimes obtained for sunrise/sunset situation by all of 3 methods. There are three reasons why such a result is obtained. First, the sunrise/sunset images are relatively rare in collected dataset as shown in Table 4 especially in Sensoji and Meiji Shrine. That is, it is difficult for small number of objects to form a cluster when applying K-means clustering. The second reason is that bad weather and lighting affected the color feature of extracted ROI. As the third reason, it is found that shooting date of some sunrise/sunset images is wrong. Such images correspond to false negatives by the time filter, which led to low precision.

Fig. 8 shows the classification results of Tokyo Tower and Mt. Fuji obtained by proposed method. It is observed the images are correctly classified into night (a), sunrise/sunset (b), cloudy (c), and shine (d) situations.

Fig. 8. Sample images of (a) night; (b) sunset/sunrise; (c) cloudy; and (d) shine situations obtained by proposed hybrid method ((1) Tokyo Tower and (2) Mt. Fuji).

Fig. 9 shows typical example images that are misclassified by content-based method but correctly classified by hybrid method. Figs. 9 (a) and (b) are cloudy situation but misclassified into night situation because the top one-third region is covered by some objects such as roof and tree. Because of light affection, Figs. 9 (c) and (d), which are night situation, are misclassified into sunrise/sunset situation. However, these misclassified images can be correctly filtered and moved to next classification stage by time windows.

(a) (b) (c) (d)

Fig. 9. Example images that are misclassified by content-based method but correctly classified by hybrid method; (a) and (b) are cloudy situation; (c) and (d) are night situation.

5. CONCLUSIONS

This paper proposed a hybrid approach for classifying sightseeing spot images to four situations. The proposed method employs ROI-based color feature extraction, and tag-based filtering in each stage. The experimental results show that the tag-based filtering can increase the accuracy, in particular precision.

In the future work, we think further improvement of tag-based filtering is possible. Furthermore, the representative image selection method will be combined with the proposed method for each situation. Considering other situations such as seasons is also challenging. The proposed method will contribute to the various kinds of tourist support services.

REFERENCES

1. J. F. Canny, "A computational approach to edge detection," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 8, 1986, pp. 679-698.
2. C. Carson, S. Belongie, H. Greenspan, and J. Malik, "Blobworld: Image segmentation using expectation-maximization and its application to image querying," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 24, 2002, pp. 1026-1038.
3. C. H. Chen, L. H. Chen, and Y. Takama, "Proposal of situation-based clustering of sightseeing spot images based on ROI-based color feature extraction," in *Proceedings of Japanese Society for Artificial Intelligence*, 2012, No. 4M1-IOS-3c-3.
4. C. H. Chen and Y. Takama, "Hybrid approach of using visual and tag information for situation-oriented clustering of sightseeing spot images," in *Proceedings of IEEE Conference on Technologies and Applications of Artificial Intelligence*, 2012, pp. 256-261.
5. A. B. Dahl and H. Aanaes, "Effective image database search via dimensionality reduction," in *Proceedings of IEEE Computer Conference on Computer Vision and Pattern Recognition*, 2008, pp. 1-6.

6. W. T. Huang, "Affinity propagation based image clustering with SIFT and color features," Master Thesis, Department of Computer Science, National Tsing Hua University, Taiwan, 2009.
7. A. K. Jain and A. Vailaya, "Image retrieval using color and shape," *Pattern Recognition*, Vol. 29, 1996, pp. 1233-1244.
8. P. A. Moellic, J. E. Haugeard, and G. Pittel, "Image clustering based on a shared nearest neighbors approach for tagged collections," in *Proceedings of International Conference on Content-based Image and Video Retrieval*, 2008, pp. 269-278.
9. N. Otsu, "A threshold selection method from gray-level histograms," *IEEE Transactions on Systems*, *Man*, *and Cybernetics*, Vol. 9, 1979, pp. 62-66.
10. B. G. Prasad, K. K. Biswas, and S. K. Gupta, "Region-based image retrieval using integrated color, shape and location index," *Computer Vision and Image Understanding*, Vol. 94, 2004, pp. 193-233.
11. S. Papadopoulos, C. Zigkolis, G. Tolias, Y. Kalantidis, P. Mylonas, Y. Kompatsiaris, and A. Vakali, "Image clustering through community detection on hybrid image similarity graphs," in *Proceedings of IEEE International Conference on Image Processing*, 2010, pp. 2353-2356.
12. A. Sleit, A. L. A. Dalhoum, M. Qatawneh, M. Al-Sharief, R. Al-Jabaly, and O. Karajeh, "Image clustering using color, texture and shape features," *KSII Transactions on Internet and Information Systems*, Vol. 5, 2011, pp. 211-227.
13. S. Silakari, M. Motwani, and M. Maheshwari, "Color image clustering using block truncation algorithm," *International Journal of Computer Science Issues*, Vol. 4, 2009, pp. 31-35.
14. A. Vailaya, M. Figueiredo, A. Jain, and H. J. Zhang, "Image classification for content-based indexing," *IEEE Transactions on Image Processing*, Vol. 10, 2001, pp. 117-130.
15. A. Vailaya, M. Figueiredo, A. Jain, and H. J. Zhang, "Content-based hierarchical classification of vacation images," in *Proceedings of IEEE International Conference on Multimedia Computing and Systems*, Vol. 1, 1999, pp. 518-523.
16. A. Vailaya, M. Figueiredo, A. Jain, and H. J. Zhang, "A Bayesian framework for semantic classification of outdoor vacation images," in *Proceedings of SPIE Storage Retrieval Image Video Databases VII*, Vol. 3656, 1999, pp. 415-426.
17. R. C. Veltkamp and M. Tanase, "A survey of content-based image retrieval systems," *Multimedia Systems and Applications Series*, Vol. 21, 2002, pp. 47-101.
18. N. Zhou, W. M. Dong, J. X. Wang, and P. Jean-Claude, "Simulating human visual perception in nighttime illumination," *Tsinghua Science and Technology*, Vol. 14, 2009, pp. 133-138.
19. D. Zhong, H. J. Zhang, and S. F. Chang, "Clustering methods for video browsing and annotation," in *Proceedings of SPIE Storage Retrieval Image Video Databases IV*, Vol. 2670, 1996, pp. 239-246.

Chia-Huang Chen (陳家煌) received the Dr. Eng. Degree from Tokyo Metropolitan University, Tokyo, Japan in 2013. He was a Research Assistant at Computer and Network Center, National Chi Nan University in Taiwan from 2009 to 2010. His current research interests include image processing, Web intelligence, and cloud computing.

Yasufumi Takama received the Dr. Eng. Degree from University of Tokyo, Tokyo, Japan in 1999. He was a JSPS (Japan Society for the Promotion of Science) Research Fellow from 1997 to 1999. From 1999 to 2002 he was a Research Associate at Interdisciplinary Graduate School of Science and Engineering, Tokyo Institute of Technology in Japan. From 2002 to 2005, he was an Associate Professor at Department of Electronic Systems and Engineering, Tokyo Metropolitan Institute of Technology, Tokyo, Japan. Since 2005, he has been an Associate Professor at Faculty of System Design, Tokyo Metropolitan University, Tokyo, Japan. He also participated in PREST (Pre-cursory Research for Embryonic Science and Technology), JST (Japan Science and Technology Corporation) from 2000 to 2003. His current research interest includes Web intelligence, information visualization, and intelligent interaction. Dr. Takama is a member of IEEE, IEICE (Institute of Electronics, Information and Communication Engineers), JSAI (Japanese Society of Artificial Intelligence), and IPSJ (Information Processing Society of Japan).