4M1-IOS-3c-3

Proposal of Situation-based Clustering of Sightseeing Spot Images based on ROI-based Color Feature Extraction

Chia-Huang Chen [*1] Lieu-Hen Chen [*2] Yasufumi Takama [*1]

[*1] Graduate School of System Design, Tokyo Metropolitan University, Japan
[*2] Department of Computer Science and Information Engineering, National Chi Nan University, Taiwan

As the amount of photos shared on album websites grows up rapidly, the meaningful grouping of these images becomes important. In particular, sightseeing spot images in different situations, such as weather conditions and seasons, are useful for tourists to decide when to visit there. This paper proposes a method to group sightseeing spot images into several different situations by a hierarchical processing based on color feature on the designed region (ROI, region of interest) and K-means clustering. At the first stage, night view images are discriminated from daytime images. Then the daytime images are divided into sunrise/sunset and other images at second stage. Finally, the cloudy images are separated from shine images in others images generated at second stage. Experimental results show that the extraction of color feature within ROI is effective for obtaining clusters with high precision and recall.

1. Introduction

The tourist usually takes a lot of photos during a journey and shares these photos on web albums such as Flickr and Picasa. 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 landmarks like towers or bridges are famous for their 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 people planning their trips.

This paper proposes a method for grouping outdoor sightseeing spot images into categories of different situations. The paper focuses on outdoor sightseeing spots because indoor spots such as museums and aquariums are less affected by situations in terms of their exhibition contents. The situations handled in this paper are night, sunrise/sunset, cloudy, and shine.

In order to establish a method that can be applied to various sightseeing spots without training data, this paper employs unsupervised learning approach. It is supposed that color has stronger relation with situation than other features. Therefore, the proposed method extracts the color feature on the designed region, based on which K-means clustering is applied. Taking into account the suitable color features for each situation, hierarchical clustering approach is employed. In the first stage, for discriminating night view images from others, the value component of HSV color space on the top 1/3 region of each image is extracted as a feature. The second stage utilizes the image segmentation for obtaining the Region of Interest (ROI). From ROI, Cb and Cr components of YCbCr space are extracted for discriminating sunrise/sunset situation from others. In the third stage, the same color features as the second stage are also employed for extracting cloudy and shine images.

Contact: Chia-Huang Chen, Graduate School of System Design, Tokyo Metropolitan University., 6-6 Asahigaoka, Hino, Tokyo, Japan., +81-42-585-8629, chen-chiahuang@sd.tmu.ac.jp

The test data set which was collected manually from Flickr contains 240 images of Mt. Fuji. The experimental results achieved the precision of 92%, 93.55%, 84.78% and 91.3% for night, sunrise/sunset, cloudy and shine situations respectively.

The rest of this paper is organized into four sections. Section 2 briefly reviews the existing works related to feature extraction and image clustering method. Section 3 describes the proposed method. In section 4, experimental results are presented to evaluate the effectiveness of our approach.

2. Related Work

Since vast amount of images are available on the Web, the effective grouping of these images into specific classes can be useful for many kinds of applications such as categorization of vocation images [Vailaya 1999 2001] and browsing of video shots [Zhong 1996]. Vailaya et al. 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 supervised learning method, clustering method can group a set of unsupervised (unlabeled) data into several clusters based on low-level visual features. Silakari et al. [Silakari 2009] 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 1000 images into 10 clusters such as busses, dinosaurs, and flowers. Sleit et al. [Sleit 2011] have utilized the histogram, Gabor filters for texture, and Fourier transformation for 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 [Huang 2009] 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 database consists of 2550 groups of 4 images each.

3. Proposed Method

As for images in a certain situation, there are specific color features in local region which are meaningful for clustering. For example, shine and sunset images have blue and orange colors in sky area respectively. Therefore, the proposed method extracts the different color features on the designated region for each situation. As features required for discriminating images of a certain situation are different from situation to situation, we divided the discrimination process into 3 stages. In each stage, based on the extracted features, the K-means method [Hartigan 1979] is applied for clustering images in terms of situation. Finally, clusters corresponding to the specific situation are discriminated by the characteristic of each cluster's centroid.

3.1 Overall Procedure

In order to achieve the purpose of grouping images into categories of different situations, the hierarchical organization of situation categories as shown in Fig. 1 is considered. After collecting sightseeing spot images from web photo sharing sites such as Flickr, most of images are supposed to be distinguished into indoor scenes, outdoor scenes, and others such as close-ups with using existing method [Vailaya 2001]. Input images of the proposed method are outdoor photos of a target sightseeing spot. The overall procedure consists of three stages considering the hierarchical structure of situation categories. At first, the outdoor images are divided into night and daytime at the first stage. The second stage separates sunrise/sunset from other images in the daytime images. At the third stage, the categories of cloudy and shine images are obtained from other images in the 2nd stage.

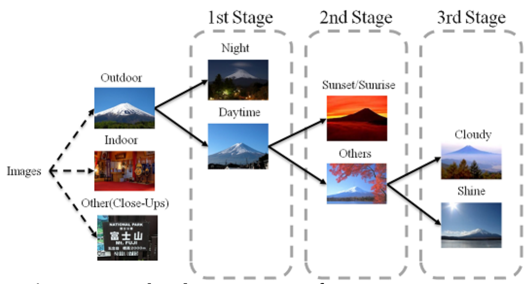

Fig. 1 Hierarchical organization of situation categories.

The processing flow as shown in Fig. 2 is applied in each stage. The local color feature is extracted at first. Then the histogram vectors of color feature are calculated as inputs to K-means clustering. Finally the situation discrimination method is applied to identify the clusters corresponding to the target situation from the obtained clusters. The procedure goes back to the block of local color feature extraction as the beginning of next stage.

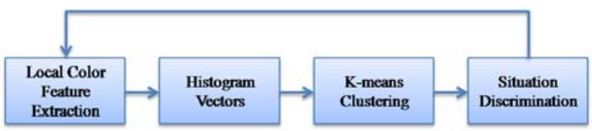

Fig. 2 Processing flow in each stage.

3.2 Processing at 1st Stage

The goal of this stage is to discriminate night images from other daytime images. In human perception, the darkness and brightness are commonly used for recognition of daytime and night [Zhou 2009]. However, the light or reflection in image will influence the result if the global brightness feature is used. Refer to the rule of thirds [Liu 2010], which is a heuristic about the composition of images, an image could be divided into 9 equal parts with two equally-spaced horizontal lines and two such vertical lines. According to our observation, the brightness within the top one-third region can contain color features enough to discriminate night images from others. In our preliminary experiment, intensity and value components of HSV were compared, and it was found the value component was better than intensity in this application. Therefore the histogram of value component within the top one-third region of an image is calculated as local color feature in the first stage. The threshold of value component is 84 (its range is [0,255]), which is determined based on the result of preliminary experiment. After thresholding, a histogram with 2 bins is calculated as input to K-means clustering. The clusters are obtained by setting K as 2. Finally the cluster with higher value in a smaller bin is considered as the night situation, as the value component of night images is almost less than the threshold value.

3.3 Processing at 2nd Stage

In this stage, a further grouping of daytime images into sunrise/sunset and other situations is considered. The ROI (region of interest) in this stage is also sky region. However, top one-third of an image usually contains other objects than sky, such as cloud and a mountain peak, which affects the color features of the region. Therefore, in order to extract sky region more exactly than the first stage, a region segmentation method using edge detection is proposed. The method consists of the following 9 steps, which are also shown in Fig. 3.

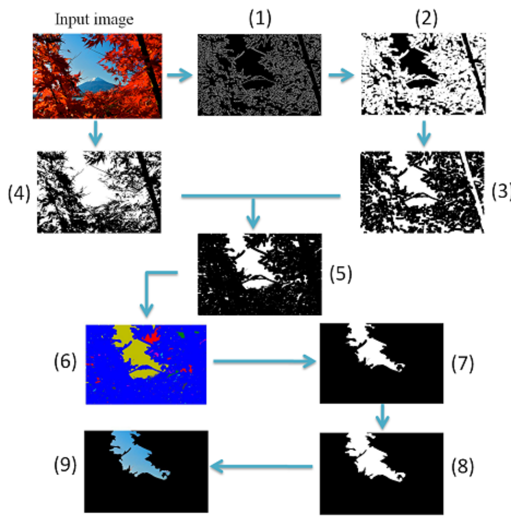

Fig. 3 Segmentation flow of ROI.

(1) Apply Canny edge detection [Canny 1986] to obtain edge region (R_e) of an input image.
(2) Dilate R_e with 5x5 kernel by morphology operation.

(3) Reverse the dilated R_e to become non-edge region (R_{ne}).
(4) Get a global image threshold of value component by Otsu's method [Otsu 1979] and perform thresholding.
(5) Get a binary image as the intersection of R_{ne} and the binary image from step (4).
(6) Apply 8-connectivity on the binary image from step (5).
(7) Get the binary image of maximal region.
(8) Obtain ROI by dilating the binary image from step (7) with 5x5 kernel.
(9) Extract the color feature within ROI.

The step (1) supposes that land region of an image contains more complicated edges or texture than the sky region. Therefore the Canny edge detection is applied to find edges. In Fig. 3 (1), white region corresponds to detected edges. By applying step (2) and (3), most of objects such as land, building, and plants can be eliminated. However, sometimes the edge detection is unavailable when the color of land region is dark. It may lead to the extraction of wrong ROI. In order to avoid such a problem, the Otsu's method is applied at step (4) to find a global image threshold of value component for eliminating the dark land region. The step (5) intersects the binary images obtained by step (3) and (4), i.e., a pixel which is white in both images is colored white in the resultant image.

In order to get each connected region, the 8-connectivity is applied at step (6). Each connected region is labeled with different values, which are indicated as different colors in Fig. 3 (6). The step (7) extracts the maximal region, which is shown as a white region a basis of ROI and colored white in Fig. 3(7). Dilation is performed on the binary image of maximal region for getting more precise ROI at step (8). Finally the color features within the ROI are extracted for clustering.

It is observed that sunset images have mostly yellow and red color in sky region, while others tend to have blue and white color in sky region. Based on this observation, the Cb and Cr components of YCbCr space within the ROI are extracted as the local color features. The number of bins and clusters are set to 128 and 8 respectively, according to the best result in a preliminary experiment.

Fig. 4 shows the examples of obtained 8 clusters including sunrise/sunset ones. The figure consists of sample images in the clusters and histograms of their centroids. 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.

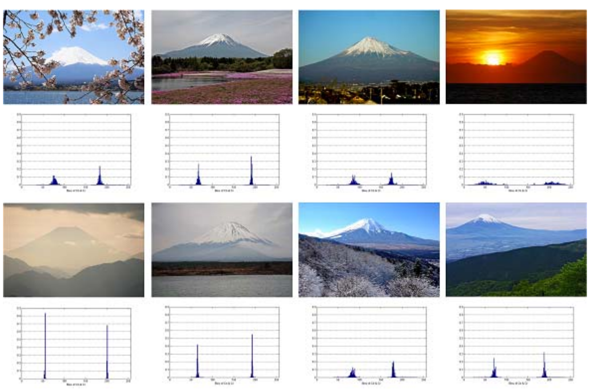

Fig. 4 Sample images in obtained clusters and histogram of their centroids at 2nd stage.

3.4 Processing at 3rd Stage

The purpose of the final stage is to group remaining images from 2nd stage into cloudy and shine situation. The ROI is segmented by the same method as the 2nd stage. After testing different combination of HSV and YCbCr space in preliminary experiments, the Cb and Cr components are employed as features also in this stage. The number of bins and clusters are set to 32 and 8 respectively, according to the best result in a preliminary experiment.

After K-means clustering, peak values of clusters' centroids in Cb component is compared with the mean value of all centroids' peak value. The cluster of which centroid has the higher peak value 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.

4. Experiments

In order to evaluate the performance of each stage as proposed in Sec. 3, the proposed method is compared with several common color spaces such as RGB, HSV, and YCbCr for examining suitable color feature for each specific situation. The effect of extracting features from ROI is also evaluated through comparison with global feature extraction.

We collected images of a sightseeing spot with Flickr, from which images corresponding to a situation (night, sunrise/sunset, cloudy and shine) were selected and labeled manually. Five people took part in the labeling process, and an image is labeled only when all of them agree to the label. Table 1 summarizes the test dataset.

Table 1 Test dataset.

Name of Spot	Night	Sunrise /Sunset	Cloudy	Shine	Total Labeled Images	Query URL
Mt. Fuji	30	30	46	134	240	http://www.flickr.com/search/?q=Mt.+Fuji&z=m

The measures of precision and recall commonly used in information retrieval are applied to evaluate the performance of the proposed method. Table 2 shows the experimental result in night situation at the 1st stage. The values of precision and recall are measured by proposed method (the value component of HSV space within top 1/3 region of image), local intensity (intensity within top 1/3 region of image), global value (the value component of HSV space within whole image) and global intensity (the intensity within whole image). It is seen that the value component performs better than intensity. It is observed that the performance of intensity gets worse when a shine image contains deep blue or dark sky. Figure 5 shows an image of Mt. Fuji 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 the recall of proposed method is lower than methods using global color feature, precision of proposed method is much higher than those methods. As the number of images of a sightseeing spot obtained from the Web is usually huge, we consider the precision is more important than recall.

Table 2 Precision and recall values of night situation at 1st stage.

Measures	Proposed	Local Intensity	Global Value	Global Intensity
Precision (%)	92.00	87.50	77.42	65.00
Recall (%)	76.67	70.00	80.00	86.67

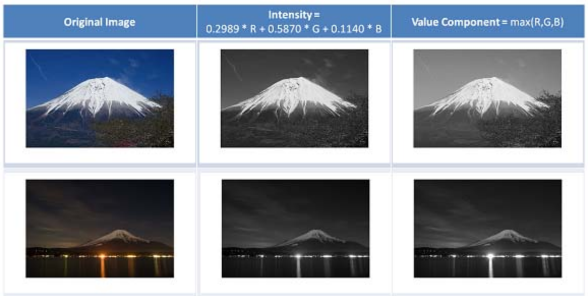

Fig. 5 The comparison of intensity and value component.

In the experiments of the 2nd and the 3rd stages, the proposed method is compared with the following methods.

- w/o-Otsu : this method is the same with proposed method but skips the step (4) as shown in Fig. 3.
- Global-CbCr : this method uses Cb and Cr components within global image.
- ROI-RGB : this method uses R, G and B components within the same ROI as the proposed method.

Table 3 shows the result of sunrise/sunset, cloudy and shine situations using proposed method and other three methods mentioned above. Obviously the proposed method can get the best results in both of precision and recall in these three situations. The result of w/o-Otsu indicates that adding the thresholding of global image with Otsu's method is effective. Comparison between the proposed method and Global-CbCr shows feature extraction from ROI is effective. The comparison of proposed method and ROI-RGB shows the Cb and Cr component is more suitable than RGB space in clustering of sunrise/sunset, cloudy and shine situations. When we compare the results of the proposed method among different situations, it is seen that the results of cloudy situation are worse than other situations. This is because the color feature extracted from cloudy image contains mostly white and gray color, whereas the Cb and Cr represent the blue-difference and red-difference chroma components, respectively.

Table 3 Precision and recall values of sunrise/sunset, cloudy and shine situations.

Situation	Measures	Proposed	w/o-Otsu	Global-CbCr	ROI-RGB
Sunrise/Sunset	Precision (%)	93.55	92.86	92.86	90.91
	Recall (%)	96.67	86.67	43.33	66.67
Cloudy	Precision (%)	84.78	81.82	82.05	76.92
	Recall (%)	84.78	78.26	69.57	43.48
Shine	Precision (%)	91.30	90.00	86.90	79.11
	Recall (%)	94.03	94.03	94.03	93.28

5. Conclusion

The situation-oriented grouping of sightseeing spot images is useful for tourists to plan which place and when to visit. In this paper, we apply a hierarchical clustering for grouping sightseeing spot images into four situations. The proposed method employs ROI-based color feature extraction, and appropriate color features are employed in each stage. The experimental results show that the proposed local color feature can get higher precision and recall than global color feature. It is also shown that the applied color space is more effective in discrimination of each situation.

In the future work, we will combine the proposed method with the representative image selection for each situation. The proposed method will contribute to the various kinds of tourist support services.

References

[Canny 1986] J. F. Canny, "A Computational Approach to Edge Detection," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 8, pp. 679-698, 1986.

[Hartigan 1979] J. Hartigan and M. Wong, "Algorithm as 136: A K-means Clustering Algorithm," Journal of the Royal Statistical Society. Series C (Applied Statistics), vol. 28, pp. 100-108, 1979.

[Huang 2009] 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.

[Liu 2010] L. G. Liu, R. J. Chen, L. Wolf, and D. Cohen-Or, "Optimizing Photo Composition," Computer Graphics Forum, vol. 29, pp. 469-478, 2010.

[Otsu 1979] N. Otsu, "A Threshold Selection Method from Gray-Level Histograms," IEEE Transactions on Systems, Man, and Cybernetics, vol. 9, pp. 62-66, 1979.

[Silakari 2009] S. Silakari, M. Motwani, and M. Maheshwari, "Color Image Clustering using Block Truncation Algorithm," International Journal of Computer Science Issues, IJCSI, vol. 4, pp. 31-35, 2009.

[Sleit 2011] 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, pp. 211-227, 2011.

[Vailaya 1999] A. Vailaya, M. Figueiredo, A. Jain, and H. J. Zhang, "Content-based hierarchical classification of vacation images," IEEE International Conference on Multimedia Computing and Systems, Vol. 1, pp. 518-523, 1999.

[Vailaya 2001] A. Vailaya, M. Figueiredo, A. Jain, and H. J. Zhang, "Image Classification for Content-Based Indexing," IEEE Transactions on Image Processing, vol. 10, pp. 117-130, 2001.

[Zhong 1996] D. Zhong, H. J. Zhang, and S. F. Chang, "Clustering methods for video browsing and annotation," in Proc. SPIE Storage Retrieval Image Video Databases IV, vol. 2670, pp. 239–246, 1996.

[Zhou 2009] N. Zhou, W. M. Dong, J. X. Wang, and P. Jean-Claude, "Simulating Human Visual Perception in Nighttime Illumination," Tsinghua Science & Technology, vol. 14, pp. 133-138, 2009.