

**Study on Situation-oriented Classification of Sightseeing  
Images Based on Visual and Metadata Features**

**September, 2013**

**Chia-Huang Chen**

**Tokyo Metropolitan University**



## Abstract

This thesis proposes a method for classifying sightseeing images into different situations based on their visual and metadata features. The widespread use of digital cameras and smart phones has brought about a situation where tourists take lots of photos of memorable moments during their travels and upload these photos to web albums such as Flickr or Picasa. These sightseeing images then become useful resources for others who plan to visit the places shown in the images. As scenes of sightseeing spots vary from situation to situation, the impression one gets from viewing these images depends heavily on conditions such as the weather and season. If a web-based tourist service could provide tourists with different views of sightseeing spots in various situations, visitors would be able to plan their vacations by looking at the views they enjoy. That is, such a service would be useful for tourists to plan when and where to visit.

To achieve this goal, a method that can classify various sightseeing images into various situations is required. Although image classification / annotation using visual and text features is becoming a major research topic in various fields, such as information retrieval and web intelligence, image classification methods focusing on various situations have not been studied yet.

One of the contributions of this thesis is to consider various situations and organize them in terms of their characteristics. The situations treated in this thesis are classified into weather-related, time-related, and season-related ones. Weather-related situations include sunshiny and cloudy situations, and color features of sky regions are expected to be effective as a means of classifying them. On the other hand, time-related situations are characterized as certain times of the day such as sunrise/sunset, daytime, and night-time. Therefore, shooting date and time, i.e., metadata attached to the photos, are important features for such a classification.

Different from weather-related and time-related situations, scenery change by season will depend on the characteristics of a sightseeing spot. It may happen that even though two sightseeing spots are geographically close, one maybe season-dependent and the other not. Therefore, sightseeing spots should also be classified into season-dependent and season-independent as a preprocessing for image classification. This thesis proposes different classification methods for each of these situation types.

The thesis consists of six chapters. Chapter 1 describes the background and motivation. The vast amount of sightseeing images available in the web albums is an important resource for tourists. The purpose of this thesis is to establish an efficient image classification method targeting sightseeing images showing various situations, which will add extra value to existing web-based tourist services. The related topics of the thesis, i.e., image classification / annotation, have attracted a lot of research, and various features and integration methods have been studied. However, the major focus of these studies has been general-purpose processing; methods focusing on various situations have not been studied yet. This chapter defines and organizes the situations to

be handled in the thesis and discusses the challenges of classifying sightseeing images into each situation.

Chapter 2 describes the existing applications of tourism informatics. Image classification and annotation methods based on supervised and unsupervised learning with various features are also covered as related work.

Chapter 3 describes content-based image classification targeting weather-related and time-related situations. Visual features for identifying each target situation are considered from viewpoints such as composition of the photos and typical colors in each situation. The images are classified in a hierarchical manner, in each stage of which efficient color features, region of interests (ROI), and cluster identification method are determined. Experimental results show that the proposed method can obtain clusters for each situation with high precision and recall.

Chapter 4 focuses on time-related situations and extends the content-based image classification method proposed in Chapter 3 by introducing filtering based on tag information. By using timestamps attached to images, clusters for the situations obtained by the content-based approach are verified to increase the accuracy of the classification. The time windows are adjusted by considering the geolocation of sightseeing spots, and this adjustment is based on information obtained from the Web. Experimental results show that this method can improve precision while maintaining recall in most cases.

Chapter 5 focuses on season-related situations and proposes a method for classifying sightseeing spots into season-dependent and season-independent ones as preprocessing for image classification. If image processing is required in order to extract features from photos, the network load for downloading photos and the cost of image processing become a serious problem. To solve this problem, the statistical features of sightseeing spots calculated using metadata are proposed. Image processing is only applied to the spots classified as season-dependent by machine learning with the statistical features. Experimental results show that this method can classify actual sightseeing spots with high precision and recall.

Chapter 6 summarizes the conclusions presented in Chapter 3 to Chapter 5. This thesis proposes three kinds of image classification methods, each of which employs efficient visual and metadata features and integration methods for the target situations. The results of this thesis are meant to contribute to tourism and related applications, which are important issues in many cities including Tokyo. As the volume of images and metadata available on the Web is still increasing at a rapid rate, the contributions of the thesis may have numerous other applications.



# Index

<b>Abstract</b> .....	i
<b>List of Figures</b> .....	v
<b>List of Tables</b> .....	viii
1. Introduction.....	1
2. Related Work .....	9
2.1. Application of Tourism Informatics.....	9
2.2. Image Classification and Annotation .....	13
2.2.1. Outline of Image Classification and Annotation .....	13
2.2.2. Machine Learning for Image Classification / Annotation .....	13
3. Visual Feature-Based Classification of Sightseeing Images into Weather-Related and Time-Related Situations.....	16
3.1. Classification of Weather-Related and Time-Related Situations .....	16
3.2. Hierarchical Organization of Image Classification Method .....	17
3.3. Processing in the First Stage .....	19
3.4. Processing in the Second Stage.....	23
3.5. Processing in the Third Stage.....	30
3.6. Experiments .....	31
4. Hybrid Approach Based on Visual and Metadata Features for Image Classification Targeting Time-Related Situations .....	40
4.1. Utilization of Tag Information for Time-Related Situations.....	40
4.2. Overall Procedure .....	41

4.3.	Definition of Time Window .....	43
4.4.	Hierarchical Classification with Time Filter .....	45
4.4.1.	Processing at the First Stage .....	45
4.4.2.	Processing at the Second Stage .....	46
4.4.3.	Processing at the Third Stage .....	48
4.5.	Experiments .....	50
5.	Identification of Season-Dependent Sightseeing Spots Based on Metadata-Derived Features and Image Processing.....	58
5.1.	Identification of Season-Dependent Sightseeing Spots .....	58
5.2.	Machine Learning in the First Phase .....	60
5.3.	Image Processing in the Second Phase .....	65
5.4.	Experiments .....	68
5.4.1.	Experimental Results of the First Phase .....	72
5.4.2.	Experimental Results of the Second Phase.....	73
6.	Conclusions and Future Research Directions .....	76
	<b>References</b> .....	78
	<b>Related Publications</b> .....	A
	<b>Acknowledgments</b> .....	B

## List of Figures

Fig. 1.1 Prototype of web-based tourist service system. (a) night-time, (b) sunrise/sunset, (c) cloudy, and (d) sunshiny situations. ....	2
Fig. 1.2 Example photos of Mt. Fuji taken in sunshiny, cloudy, night-time, and sunrise/sunset situations. ....	4
Fig. 1.3 Photos of Ueno Park taken in 2011. ....	5
Fig. 1.4 Photos of Roppongi Hills taken in 2011. ....	6
Fig. 1.5 The organization of the thesis. ....	6
Fig. 2.1 Check-in places shown on map of Facebook. ....	10
Fig. 2.2 Geotagged photos shown on World Map of Flickr. ....	11
Fig. 2.3 EXIF data available on Flickr. ....	11
Fig. 2.4 Suggestions for sights shown on foursquare. ....	12
Fig. 3.1 Hierarchical organization of situation categories. ....	17
Fig. 3.2 Processing flow in each stage. ....	18
Fig. 3.3 Application of the rule of thirds to images in daytime and night-time situations. .	19
Fig. 3.4 Comparison of intensity and value components. ....	20
Fig. 3.5 Color histograms of value component calculated from the top 1/3 region of example images in daytime and night-time situations. ....	21
Fig. 3.6 Illustration of situation discrimination in the first stage. ....	22
Fig. 3.7 Segmentation flow of ROI and example diagrams of histogram and K-means. ....	24
Fig. 3.8 Sample images in obtained cluster 1 to 4 and histograms of Cb & Cr values of their centroids in the second stage. ....	27
Fig. 3.9 Sample images in obtained cluster 5 to 8 and histograms of Cb & Cr values of their centroids in the second stage. ....	28

Fig. 3.10 Images of individual situations for Mt. Fuji. (a) night-time and (b) other situations in the first stage. (c) sunrise/sunset situation and (d) daytime in the second stage. (e) cloudy and (f) sunny situations in the third stage. ....	39
Fig. 4.1 Overall processing flow.....	42
Fig. 4.2 Example of cloudy images that are mis-classified by content-based method but correctly classified by hybrid method. ....	45
Fig. 4.3 Sample images in obtained clusters and histogram of their centroids at second stage. ....	47
Fig. 4.4 Example of night-time images that are mis-classified by content-based method but correctly filtered and removed by hybrid method.....	48
Fig. 4.5 Collected and filtered region of Mt. Fuji. ....	51
Fig. 4.6 System architecture for sightseeing images classification and rendering in different situations. ....	52
Fig. 4.7 Sample images (top row) that are mis-classified into sunrise/sunset cluster and their ROI (bottom row). The first and second image is labeled as cloudy and night situation respectively.....	56
Fig. 4.8 Sample images of (a) night-time, (b) sunset/sunrise, (c) cloudy, and (d) sunny situations obtained by proposed hybrid method ( (1) Tokyo Tower and (2) Mt. Fuji).....	57
Fig. 5.1 Example photos of Shinjuku Gyoen and Rainbow Bridge Tokyo taken in April, July, November, and January.....	58
Fig. 5.2 Overall processing flow.....	59
Fig. 5.3 Monthly number of tourists of Shinjuku Gyoen and Rainbow Bridge Tokyo in 2011.....	60
Fig. 5.4 Illustration of attributes 1 and 7 for Shinjuku Gyoen and Rainbow Bridge Tokyo in	

2009, 2010 and 2011.....	62
Fig. 5.5 Illustration of attributes 2-6 for Shinjuku Gyoen during 2009 to 2011.....	63
Fig. 5.6 Image processing flow in the second phase. ....	67
Fig. 5.7 Photos of Ueno Park collected from Flickr. ....	69
Fig. 5.8 Photos of Roppongi Hills collected from Flickr. ....	70

## List of Tables

Table 1.1 Types of situations. ....	4
Table 3.1 Precision and recall values of night-time situation experimented on different threshold of value component. ....	21
Table 3.2 Precision and recall values of sunrise/sunset situation using RGB components.	26
Table 3.3 Precision and recall values of sunrise/sunset situation using CbCr components.	26
Table 3.4 Precision and recall values of cloudy situation using saturation component. ....	30
Table 3.5 Precision and recall values of cloudy situation using CbCr components. ....	30
Table 3.6 Test dataset summary. ....	31
Table 3.7 Precision and recall values of night-time situation in the first stage. ....	32
Table 3.8 Precision and recall values of sunrise/sunset situations in the second stage. ....	33
Table 3.9 Precision and recall values of cloudy situations in the third stage. ....	33
Table 3.10 Precision and recall values of sunshiny situations in the third stage. ....	34
Table 3.11 Best value, average, and standard deviation for precision and recall. ....	35
Table 3.12 Comparison of proposed method with baseline method in average of precision and recall. ....	36
Table 3.13 Comparison of proposed method with baseline method in average of F-measure. .....	36
Table 3.14 Total number of images and computational costs of proposed method in average and variance. ....	38
Table 4.1 Sun rising and setting times in the first day of each month at Tokyo, Shizuoka, and Kyoto. ....	43
Table 4.2 Time window as filters for different situations in April. ....	44
Table 4.3 Parameter settings of image collection. ....	50

Table 4.4 Test dataset.....	52
Table 4.5 Average precision and recall values (%) measured by different methods in each situation.....	53
Table 4.6 Comparison of hybrid method and content-based method in F-measure (%). ....	55
Table 5.1 Summary of test dataset.....	71
Table 5.2 Comparison of different attribute sets (without discretization) applied on 4 classifiers.....	72
Table 5.3 Comparison of different attribute sets (with discretization) applied on 4 classifiers.....	73
Table 5.4 Comparison of different image processing method.....	74
Table 5.5 Performace of proposed method including 1st and 2nd phase. ....	74





# 1. Introduction

Tourism informatics has been one of hot research topics with rapid development of Web technology [22], [23]. Travelers get used to looking for the information of sightseeing spots on official website or online forum for planning. Recently, web albums such as Flickr<sup>1</sup> and Picasa<sup>2</sup> 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 heavily 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.

Classification of sightseeing images into different situations is expected to add extra value to existing web-based tourist service. For example, Fig. 1.1 shows a screenshot of the prototype tourist service that maps only images of a certain situation, which are obtained by the method proposed in the thesis, based on geotag information. This kind of 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. Related works include image classification / annotation [6-12, 20, 21, 39-42], and retrieval [1-5]. There are many effective methods for image retrieval, some of which are applied to web image searches [1-5]. The main purpose of image retrieval is to find similar images with a given (query) image. Image classification / annotation aims to identify concepts / classes related with a given image. In the case of image annotation, multiple concepts can be assigned to a single image. The ImageCLEF [39], which is an evaluation forum for the cross-language annotation and retrieval of images, organizes a task of photo annotation and retrieval. In ImageCLEF 2012 [40], 94 concepts are used for photo annotation task.

Various visual features have been studied for these kinds of applications, such as color, shape, texture, and SIFT (Scale-Invariant Feature Transformation) [35]. Furthermore, be-

---

<sup>1</sup> <http://www.flickr.com/>

<sup>2</sup> <http://picasa.google.com/>

cause of spreading 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. These kinds of text information are becoming one of important features of images in addition to visual features.

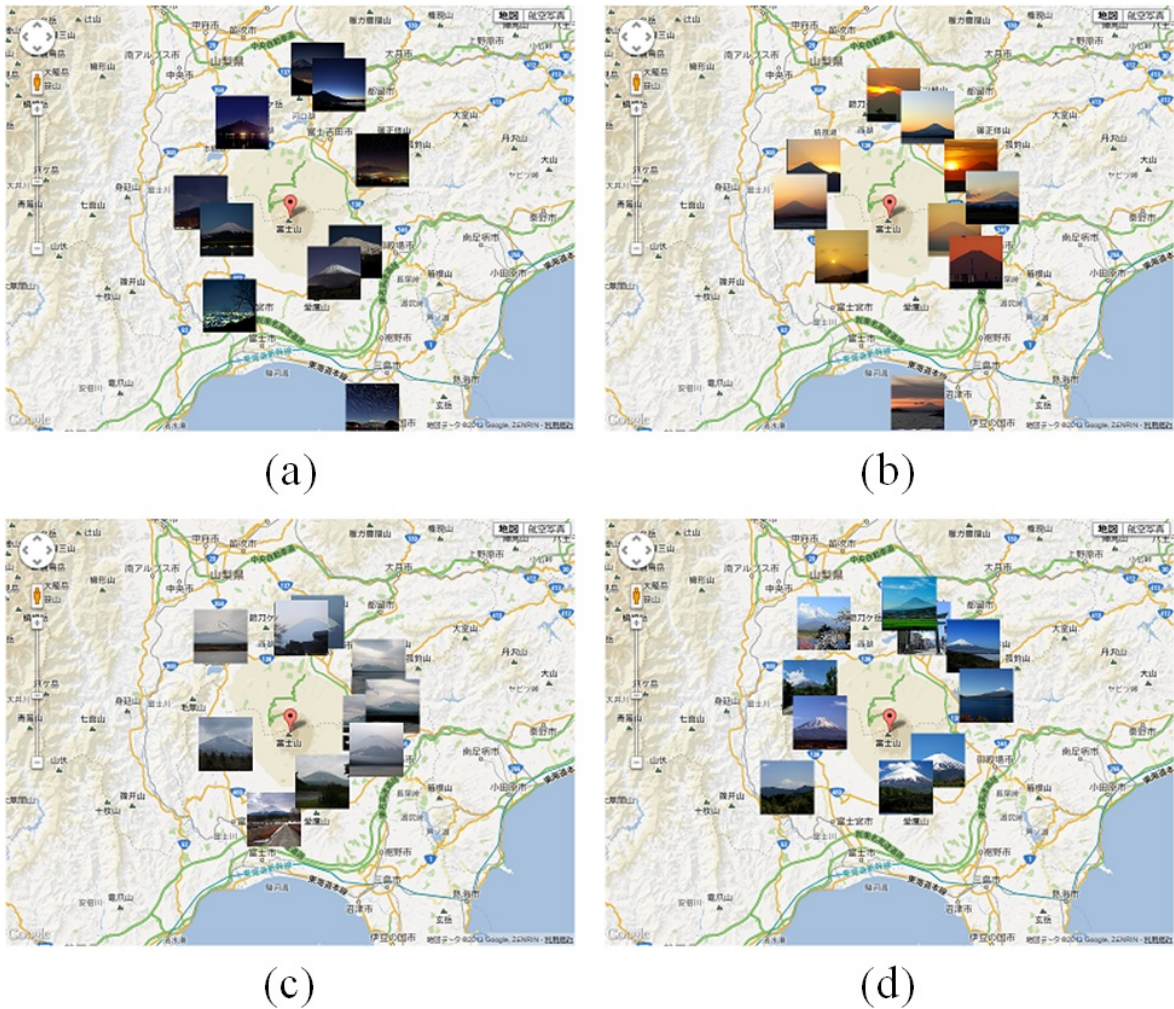


Fig. 1.1 Prototype of web-based tourist service system. (a) night-time, (b) sunrise/sunset, (c) cloudy, and (d) sunny situations.

As above-mentioned, related topics of the thesis have recently been attracting many researchers in the world, and various features and integration methods of these features have been studied. However, major focuses of those studies have been on general-purpose processing. When we consider the application of web tourist services, efficient methods focusing on various types of situations should be established. However, related technologies including organization of various situations, effective features for situation discrimi-

nation, and integration methods of those features have not been studied yet.

This thesis proposes a method for grouping outdoor sightseeing images into categories of different situations. The proposed method focuses on outdoor sightseeing spots because indoor spots such as museums and aquariums are less affected by situations in terms of their exhibition contents.

As we mention above, the view of sightseeing spot changes according to time and weather conditions. Because people can see physical objects and different color because the light would be reflected or refracted, the light source is an important factor for the variation of scene, no matter what it is natural or artificial. As sunlight changes according to the movement of the sun, a unit of time should be taken into account. However, the photos taken within few hours in the same place have not changed a lot because sun moves very slowly. People can feel the view has changed when the movement causes drastic change of sunlight, such as sunrise and sunset. As a result, it is enough to divide a day into night, sunrise/sunset, and daytime for the purpose of the thesis. On the other hand, the variation in sequence of days does not need to consider in the thesis, because the view would be similar within the same period of hours. Therefore, the situations of night-time, sunrise/sunset, and daytime are taken into account as time-related situations.

Regarding longer-term variation, only the seasonal variation will influence the sightseeing images because the color of natural objects such as flowers, grasses, and trees has changed. Although scenery would change over the years, such change is neither periodical nor repeatable. That is, photos taken in the past and are totally different from current state cannot provide valuable information for tourists who want to determine when to visit there. Therefore, only the season-related situation is considered in the thesis as long-term variation.

There is another kind of situation, weather, which could cause the change of view in non-periodic but repeatable manner. Weather can be roughly divided into sunshiny, cloudy, rainy, and snowy. The rainy situation is not suitable for sightseeing outside and taking photos, but the snow view is attractive. Because of the snow view appears in winter it is considered in seasonal situation. Therefore, it is enough to consider the sunshiny and cloudy situations as weather-related situation.

Based on these considerations, Situations handled in this thesis are night-time, sunrise/sunset, cloudy, sunshiny, and season. This thesis organizes these situations in terms of the characteristics as shown in Table 1.1. Situations are classified into 3 types: weath-

er-related, time-related, and season-related ones.

Table 1.1 Types of situations.

Type	Situations
Weather-related	Sunshiny, Cloudy
Time-related	Night-time, Sunrise/Sunset, Daytime
Season-related	Spring, Summer, Autumn, Winter



**Sunshiny**



**Cloudy**



**Night-time**



**Sunrise/Sunset**

Fig. 1.2 Example photos of Mt. Fuji taken in sunshiny, cloudy, night-time, and sunrise/sunset situations.

Fig. 1.2 shows example photos of sunshiny, cloudy, night-time, and sunrise/sunset situations. Those situations correspond to weather-related and time-related situations. Weather-related situations include sunshiny and cloudy situations. In order to discriminate sunshiny images from cloudy ones, color features of sky region are expected to be effective. On the other hand, important characteristic of time-related situations such as sunrise/sunset,



daytime, and night-time is that each situation occurs at certain times of the day. Therefore, shooting date and time, which is one of metadata attached to photos, is also important feature in addition to color features.

Different from weather-related and time-related situations, scenery change by season-related situations will depend on the characteristics of a sightseeing spot. Even though two sightseeing spots are geographically close, one maybe season-dependent and the other not. For example, Ueno Park and Mt. Takao in Tokyo, and Kinkakuji in Kyoto are season-dependent, of which sceneries change according to colors of trees or flowers. On the other hand, the Rainbow Bridge, Roppongi Hills, and Disneyland, of which main objects are buildings, are examples of season-independent. Fig. 1.3 and Fig. 1.4 show the photos that were taken in January, April, July, and October 2011 of Ueno Park and Roppongi Hills respectively. Therefore, classifying sightseeing spots into season-dependent and season-independent ones is required as preprocessing of image classification.



Fig. 1.3 Photos of Ueno Park taken in 2011.



Fig. 1.4 Photos of Roppongi Hills taken in 2011.

Considering the different characteristics of these 3 situation types, this thesis proposes different classification methods for each type of situations. The thesis consists of 6 chapters. Fig. 1.5 shows the organization of the thesis.

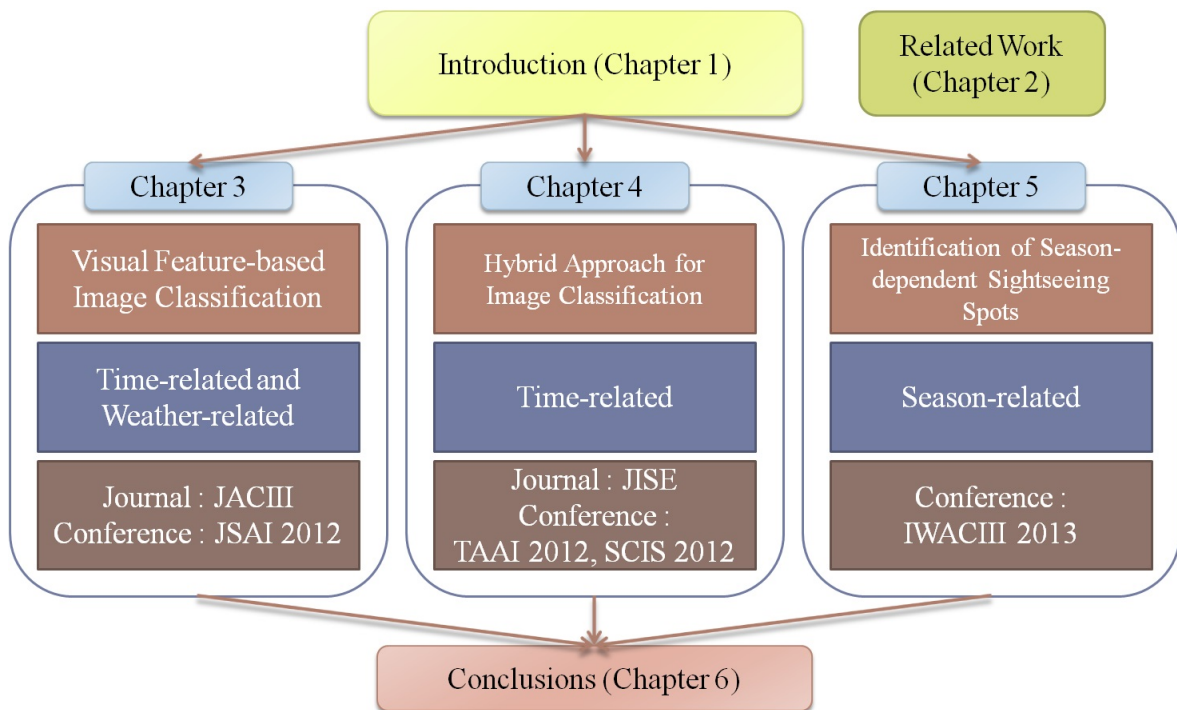


Fig. 1.5 The organization of the thesis.

Chapter 2 introduces existing applications of tourism informatics. Image classification / annotation methods based on supervised / unsupervised learning with various features are also introduced as related work.

Chapter 3 describes content-based image classification method targeting weather-related and time-related situations. Visual features for identifying each target situation are considered from viewpoints such as composition of the photos and typical colors in each situation. The images are classified in a hierarchical manner, in each stage of which efficient color features, region of interests (ROI), and cluster identification method are determined. Total 2373 images of 7 sightseeing spots on Flickr are collected, from which test data set are generated by manually classifying those images into target situations. Experimental results show that the proposed method can obtain precision of 99.28% and recall of 98.34% respectively in the best case.

Chapter 4 focuses on time-related situations, and extends content-based image classification method proposed in Chapter 3 by introducing the filtering based on tag information of images. By using timestamps attached to images, clusters for the situations obtained by content-based approach are further verified to increase the accuracy of the classification. The time windows are adjusted by considering the geolocation of sightseeing spots, and this adjustment is based on information obtained from the Web. Total 15837 images of 7 sightseeing spots are collected by setting a bounding box together with their geotag and shooting timestamps. After removing irrelevant images and giving labels of situations manually to remaining images, a data set containing 4412 images is generated. Experimental results show that the proposed method can improve precision while maintaining recall in most cases.

Chapter 5 focuses on season-related situations, and proposes a method for classifying sightseeing spots into season-dependent and season-independent ones as preprocessing for image classification. The proposed method employs machine learning approach, which requires many photos as training data. If image processing is required in order to extract features from photos, the network load for downloading photos and the cost of image processing become a serious problem. To solve this problem, the statistical features of sightseeing spots calculated using metadata are proposed. Two-stage classification approach is also proposed, which consists of light-weight classification without actual images at the first stage and color-based classification applied to only the spots classified as season-dependent at the first stage. The experimental results show the proposed method

achieved precision of 75.9% and recall of 73.3% in a test set containing 80 sightseeing spots.

Chapter 6 summarizes conclusions presented in Chapter 3 to Chapter 5, and discusses the contribution of the thesis.



## 2. Related Work

### 2.1. Application of Tourism Informatics

Since the 1980s, Information Communication Technologies (ICTs) have been bringing change to tourism globally [22]. The rapid growth in the number of online users as well as the increasing rate of online transactions can be clear evidences of the popularity of the technology [23]. With the development of search engines, high transportation speed of networks, and wide spread of digital cameras and smart phones, the convenience of planning and traveling experience is greatly improved for the tourists around the world. These technologies also make the tourism a big market having one of the biggest users of Web technologies. As a result, many innovative ideas have been constantly applied to the tourism [36]. Sharda [36] introduced three important aspects of tourism informatics which include travel recommender systems, social communities, and user interface design. Various kinds of research were reviewed within these aspects such as knowledge-based travel advisor systems, social networking for generating travel ideas, map-based interface, and Web 2.0 tourism sites.

Nowadays, tourism system provides not only the information of sightseeing spots but also personalized recommendations. Lucas et al. [24] have proposed a recommendation method and implemented a tourist recommender system. A clustering method is applied to classify users into several groups by using attributes such as age, postal code, level of education, time spent on seeing an item, and number of mouse clicks. However, this kind of suggestion doesn't take the visual attraction into account, despite the fact that sightseeing photos taken in different time can influence tourists' decision on when to visit there.

In a relatively short time Facebook<sup>3</sup> has become a major social network service. White [37] has investigated the social aspects of tourism informatics based on the travel photographs posted on the Facebook. He explored how the photos which are taken, displayed and recorded on Facebook reinforce the travel experience for tourists. It was also investigated that how these images influence the travel decisions of those who view the photos. The check-in function of Facebook provides important record of visited places.

---

<sup>3</sup> <https://www.facebook.com/>

With the record, useful information about sightseeing spots such as photos, address, website, or Email address, can be shown on the map of Facebook. It can guide the tourists who never visit there and help them to find other interesting places. Fig. 2.1 shows the counting of check-in on map of Facebook.

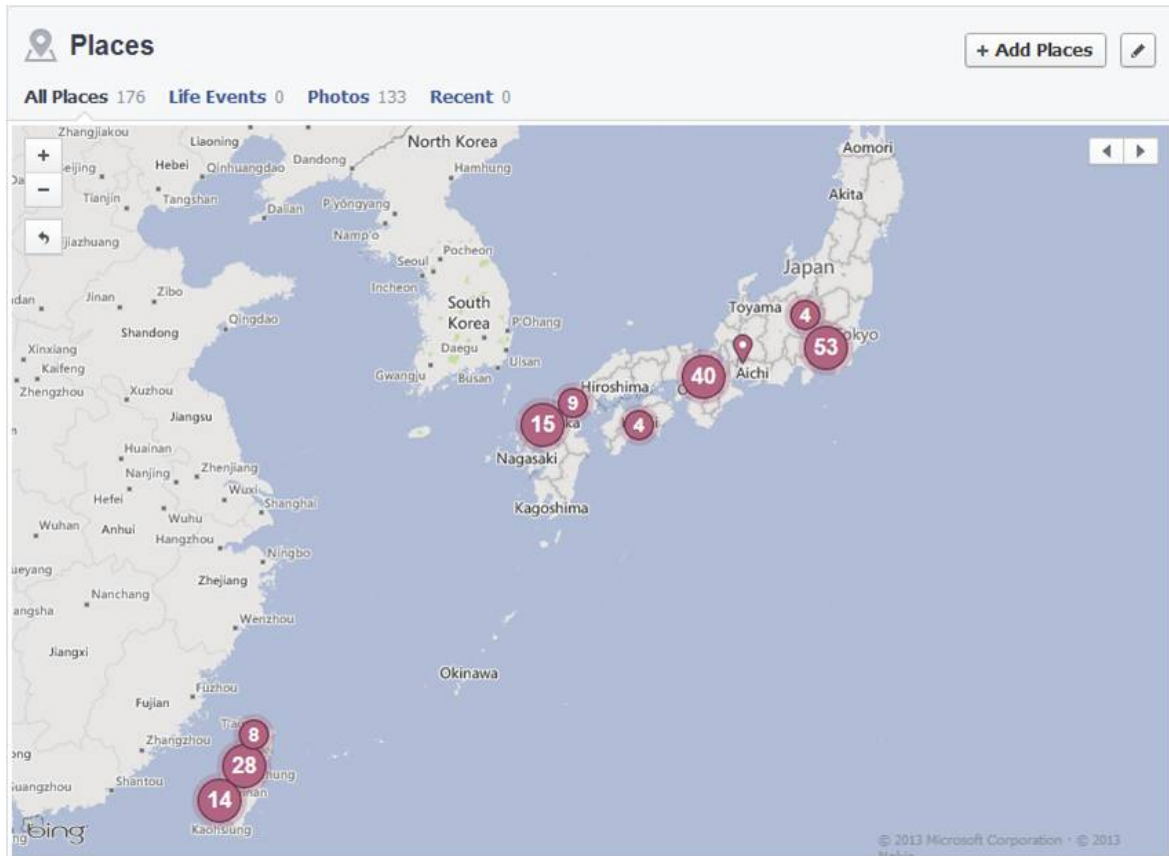


Fig. 2.1 Check-in places shown on map of Facebook.

Flickr provides online photo management and sharing application in the world. It also provides the World Map application, which can be used for searching and exploring geotagged photos with a map-based interface as shown on Fig. 2.2. Useful metadata such as EXIF (Exchangeable Image File Format) are also available on Flickr. Fig. 2.3 shows the EXIF data of photo on Flickr, which includes the information of shooting date, camera, ISO speed, resolution, latitude, longitude, etc.

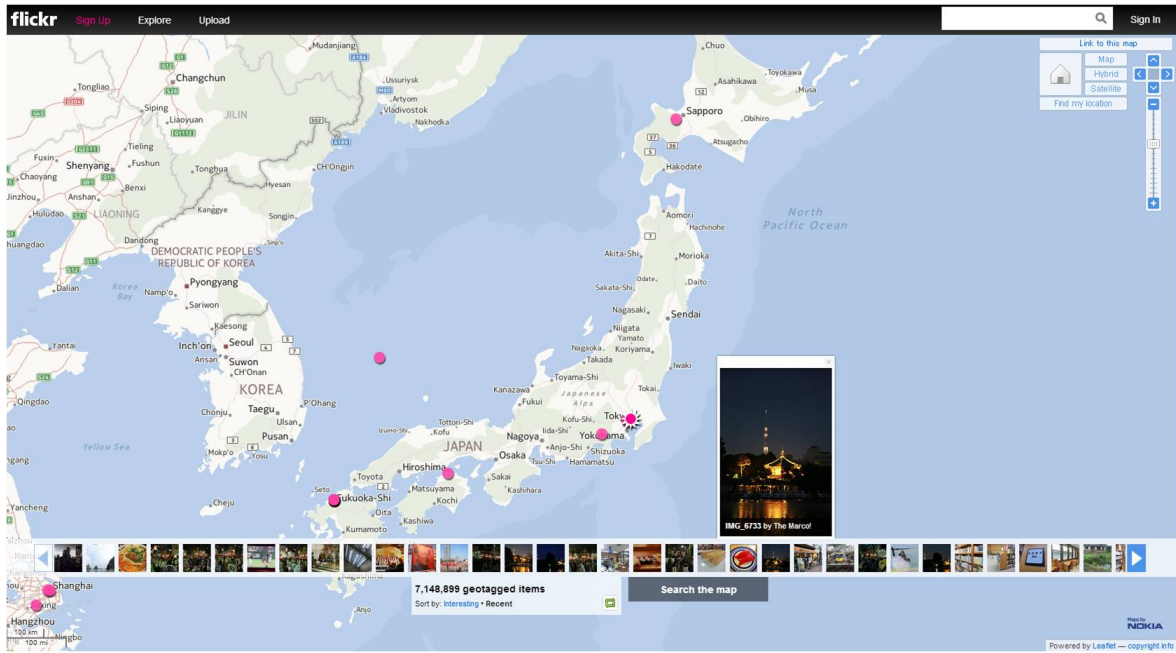


Fig. 2.2 Geotagged photos shown on World Map of Flickr.

## Photo / Exif



### What is Exif data?

Exif data is a record of the settings a camera used to take a photo or video. This information is embedded into the files the camera saves, and we read and display it here.

### Dates

Taken on	September 30, 2011 at 7.25PM PDT
Posted to Flickr	October 17, 2011 at 9.18AM PDT

### Exif data

Camera	Fujifilm FinePix S5600
Exposure	1
Aperture	f/5.6
Focal Length	9.8 mm
ISO Speed	200
Exposure Bias	0 EV
Flash	Off, Did not fire
X-Resolution	72 dpi
Y-Resolution	72 dpi
Orientation	Horizontal (normal)
Software	Adobe Photoshop Elements 5.0 (20060914.r.77) Windows
Date and Time (Modified)	2011:10:17 23:01:54
YCbCr Positioning	Co-sited
Exposure Program	Manual
Date and Time (Original)	2011:09:30 19:25:29+07:00
Date and Time (Digitized)	2011:09:30 19:25:29

Fig. 2.3 EXIF data available on Flickr.

A tourist service developed as both of a website and an application of smart phone, which is called foursquare<sup>4</sup>, is becoming popular recent year. It helps users to share and save the places they visit. Moreover, it provides personalized recommendation to inspire users when they are looking for what to do or where to go. Fig. 2.4 shows several suggestions for sights, which are retrieved according to user's location. The photo and information of sights are displayed on the left side with corresponding label number tagged on the map.

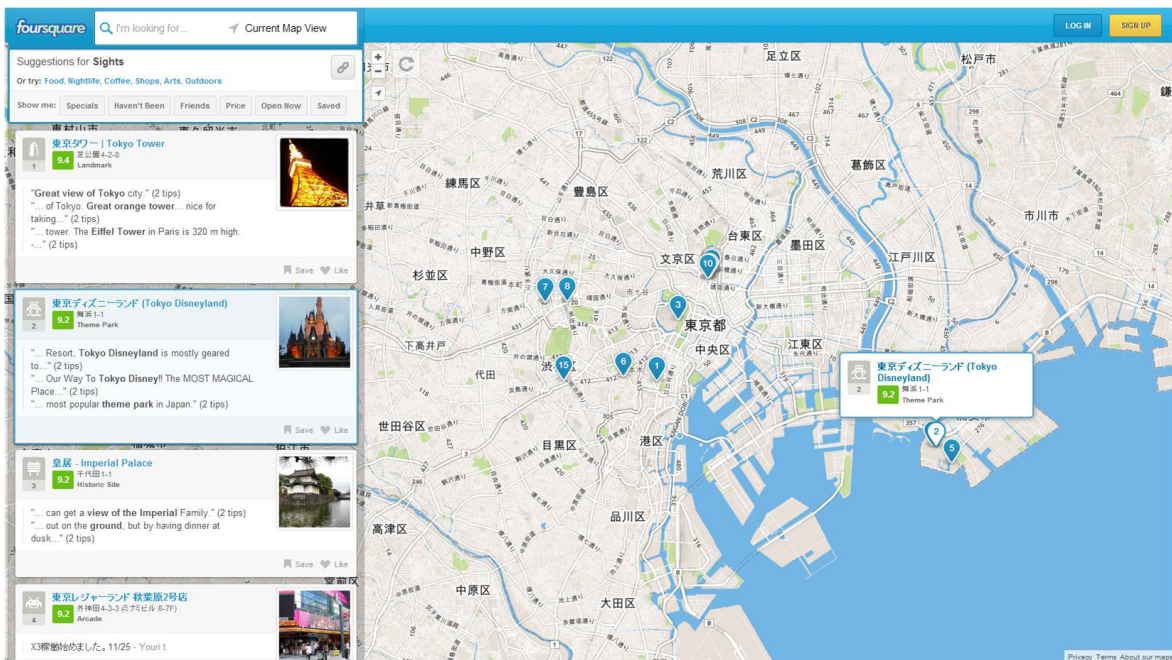


Fig. 2.4 Suggestions for sights shown on foursquare.

<sup>4</sup> <https://foursquare.com/>

## **2.2. Image Classification and Annotation**

### **2.2.1. Outline of Image Classification and Annotation**

Since the amount of available images on the Web grows up rapidly, effectively grouping these vast numbers of images into meaningful classes is useful in many kinds of applications, such as indexing of image databases [6], categorization of traveling images [7, 8], classification or mining for large-scale image collections [29, 30], and browsing of video shots [9]. Generating or selection of representative image employs image classification algorithm in process as well [31-34].

ImageCLEF [39] provides an evaluation forum for the cross-language annotation and retrieval of images. Among various tasks provided by ImageCLEF, those relating with image classification is photo annotation tasks. Photo annotation aims to identify concepts related with a given image. Different from image classification that assigns one single class for each images, multiple concepts (classes) can be assigned to a single image. Total 94 concepts such as time of day (day, night, sunrise/sunset), combustion (fire, smoke, fireworks), and flora (tree, plant, flower, grass) were used in imageCLEF 2012 [40]. The image set was annotated with 99 concepts such as the scene (indoor, outdoor, landscape), depicted objects (car, animal, person), and the representation of image content (portrait, graffiti, art) in imageCLEF 2011 [41]. In imageCLEF 2010 [42], total number of used concepts were 93, which include abstract categories such as Family&Friends or Partylife, the time of day (day, night, sunny), and quality (blurred, underexposed).

### **2.2.2. Machine Learning for Image Classification / Annotation**

Various machine learning methods including supervised and unsupervised learning have been applied to image classification. Vailaya et al. [6-8] attempted to use binary Bayesian classifiers for capturing high-level concepts from low-level image features. This method employs the approach of hierarchical classification of vacation images. Images are classified into indoor and outdoor classes on the first level, and outdoor images are further grouped into city and landscape classes. Landscape images are finally classified into sunset, forest, and mountain classes.

In order to classify a much larger dataset of images, Li et al. [29] used multiclass support vector machines to learn models for various classification tasks from labeled da-

taset of nearly two million images. The visual features obtained by clustering local interest point descriptors into a visual vocabulary are employed as descriptors. Textual tags attached to photos by Flickr users are also considered as additional features.

Quack et al. [30] proposed a mining method of objects and events. The retrieved photos are clustered according to the visual features of SURF (Speeded Up Robust Features) [38] and text features including tags, title, and description of photos. The resulting clusters are analyzed and automatically classified into objects and events.

The image classification is also employed as a process for selection of representative image. Zhou et al. [31] proposed to use visual context learning to discover visual word significance and developed Weighted Set Coverage algorithm to select canonical images containing distinctive visual words.

The canonical views for a tourist attraction should be representative of the site and exhibit a diverse set of views. Therefore, Yang et al. [32] employed visual features to encode the content of photographs and to infer the popularity of each photograph. After the encoding and the inference, they ranked photographs using a suppression scheme to keep popular views top-ranked while demoting duplicate views. After generating the ranking, canonical views at various granularities can be retrieved in real-time.

Kennedy and Naaman [33] used unsupervised methods to extract representative views and images for each landmark. The location and other metadata, as well as tags associated with images, and visual features of images are used to generate representative sets of images.

Unlike supervised learning methods, clustering methods group sets of unsupervised (unlabeled) data into clusters based on low-level visual features. Silakari et al. [10] focused on color feature of images. The color moment and Block Truncation Coding (BTC) are used to extract features and a K-means clustering algorithm is applied to group 1000 images into 10 clusters such as busses, dinosaurs, and flowers.

Sleit et al. [11] utilized color histograms, Gabor filters, and Fourier transformation for color, texture, and shape feature extraction, respectively. Based on these features, images are classified based on K-means clustering. The resultant image database included four different groups : dinosaurs, flowers, busses and elephants.

Huang [12] integrated a local SIFT (Scale Invariant Feature Transformation) feature with a global CLD (Color Layout Descriptor) feature and adopted an affinity propagation clustering algorithm that does not need to initialize the number of clusters. The bag of vis-

ual word model is applied in e-clustering to enhance clustering performance. The dataset consists of 750 groups, each of which contains 4 images.

In order to establish a clustering technique that can handle the massive amounts of user-generated photos, Papadopoulos et al. [20] 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.

Moellic, et al. [21] 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, which are classified in ten well separated categories.

### **3. Visual Feature-Based Classification of Sightseeing Images into Weather-Related and Time-Related Situations**

#### **3.1. Classification of Weather-Related and Time-Related Situations**

This chapter proposes image classification method targeting weather-related and time-related situations. Target situations include cloudy and sunshiny as weather-related situations and night-time, sunrise/sunset, daytime (cloudy and sunshiny) as time-related situations. Fig. 1.2 shows example images of each target situation.

As noted in Chapter 1, the characteristic of weather-related situations is color features within sky region. Therefore, local color features extracted from sky region is expected to be useful. Although time-related situations has different characteristics from weather-related ones, the same color features is expected to be useful as well.

Based on these considerations, the proposed method in this chapter employs visual features for classifying sightseeing images into night-time, daytime, sunrise/sunset, cloudy, and sunshiny situations. One of the challenges the proposed method should solve is how to determinate the sky region of an image exactly enough to achieve accurate classification. Another challenge is how to identify the situation after separating images according to color features. Preliminary analysis on sightseeing images has revealed that efficient extraction methods of sky region and criteria for the identification differ from each situation. Therefore, proposed method classifies images in hierarchical manner, in each stage of which efficient color features, region of interests (ROI), and cluster identification method are determined.



### 3.2. Hierarchical Organization of Image Classification Method

As noted in Sec. 3.1, color features are the most basic information for weather-related and time-related situations and are easy to use for measuring similarity between images. Global color features contain many noise, however, that recognition becomes very difficult. For images in certain situations, there are specific color features in local regions that are meaningful in clustering. As shown in Fig. 1.2, sunshiny and sunset, for example, images have blue and orange colors in the sky area, respectively, so the proposed method extracts different color features in the region designated for each situation. Features required for discriminating images in a certain situation are different from situation to situation, so we divided the discrimination process into 3 stages. In each stage, based on extracted features, the K-means method [13, 14] is applied for clustering images by the situation. Clusters corresponding to specific situations are finally discriminated by the characteristic of each clusters centroid.

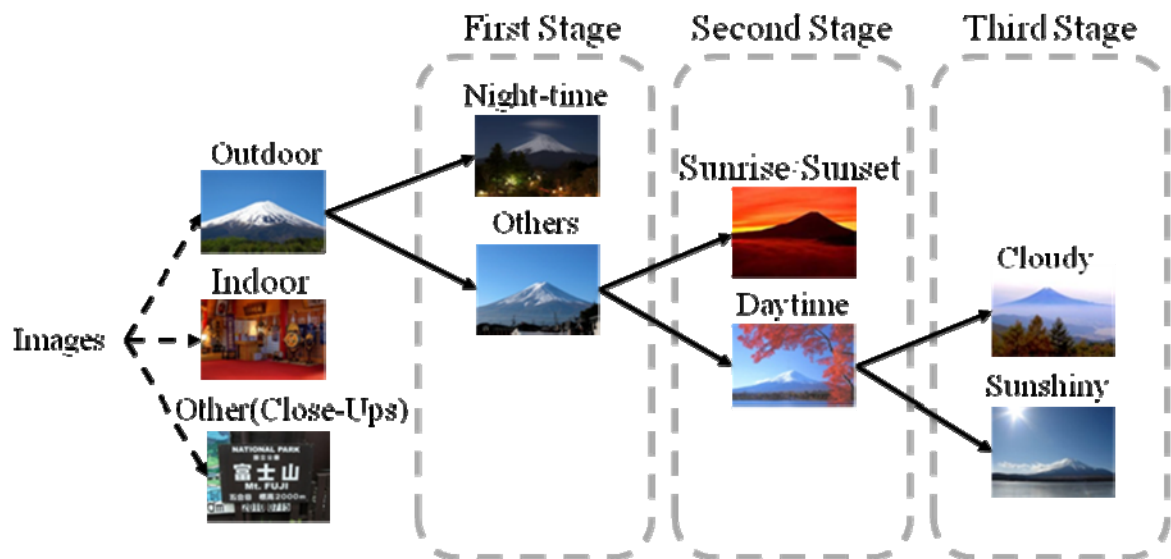


Fig. 3.1 Hierarchical organization of situation categories.

The hierarchical organization of situation categories as shown in Fig. 3.1 is considered in order to achieve the purpose of grouping images into categories for different situations. After sightseeing images are collected from photo sharing websites such as Flickr, most images are supposed to be distinguished into indoor scenes, outdoor scenes, and others such as close-ups by using existing methods [6]. Input images for 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. First, outdoor images are divided into night-time and daytime in the first stage. The second stage separates sunrise/sunset from other images among daytime images. In the third stage, categories of cloudy and sunny images are obtained from other images in the second stage.

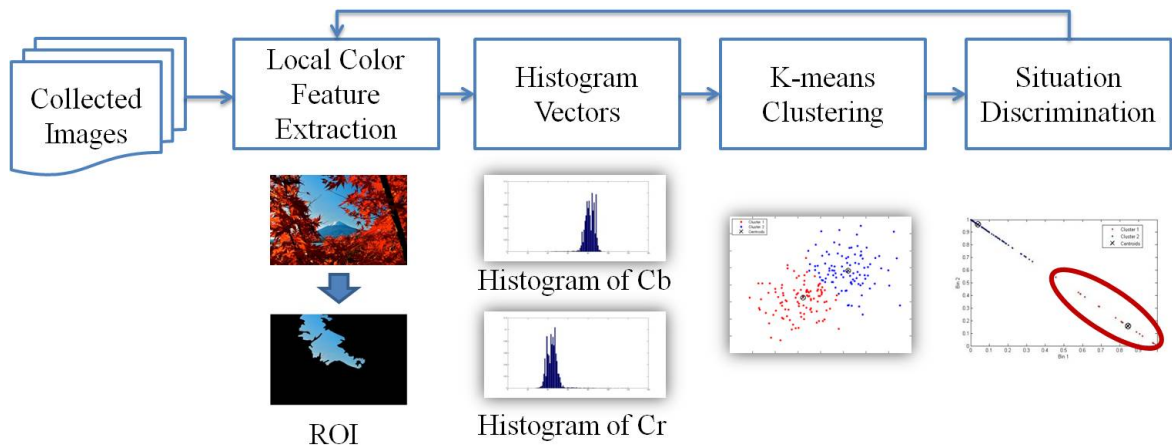


Fig. 3.2 Processing flow in each stage.

The processing flow, as shown in Fig. 3.2, is applied in each stage. Some examples of image and diagram are also shown for illustration. A local color feature is first extracted. Then histogram vectors of color feature are calculated as input to K-means clustering. The situation discrimination method is then applied to identify clusters corresponding to target situation from obtained clusters. The procedure goes back to the block of local color feature extraction that corresponds to the beginning of the next stage.

### 3.3. Processing in the First Stage

The goal of the first stage is to discriminate night-time images from others images. In human perception, darkness and brightness are commonly used for the recognition of day-time and night-time [15], so brightness is useful for discriminating night-time images from others. Light or reflection in images influences results, however, if a global brightness feature is used.

The rule of thirds [16, 17], which is a heuristic related to the composition of images, indicates that an image should be divided into 9 equal parts with two equally spaced horizontal lines and two equally spaced vertical lines. Fig. 3.3 shows images of Mt. Fuji in daytime and night-time situations, each of which is divided with the principle of the rule of thirds. Important compositional elements should be placed along these lines or located at their intersections. A “beautiful” picture is supposed to satisfy this rule to some extent.

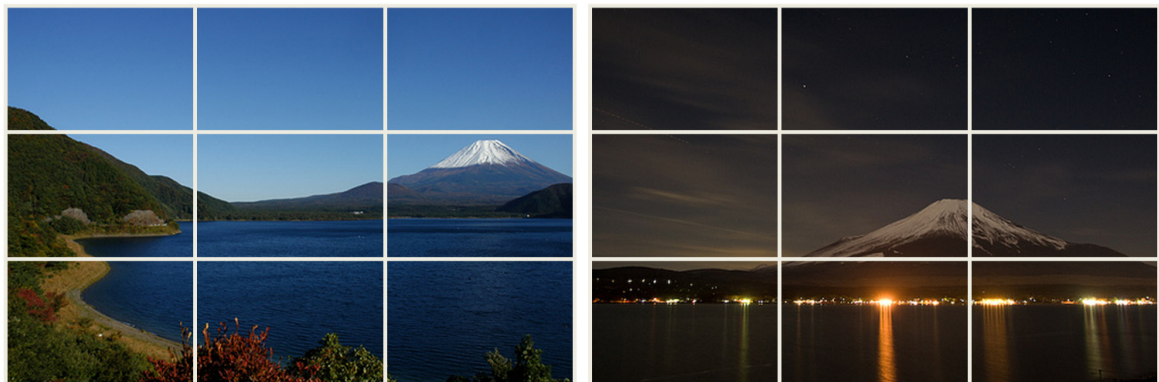


Fig. 3.3 Application of the rule of thirds to images in daytime and night-time situations.

According to our observation, brightness within the top one-third of a region contains enough color features to discriminate night-time images from others. As shown in Fig. 3.3, the top one-third of region in daytime has relatively high brightness than that in night-time, but other region does not because it would be covered by shadow or dark ground.







Original Image	Intensity = $0.2989 * R + 0.5870 * G + 0.1140 * B$	Value Component = $\max(R,G,B)$
		
		

Fig. 3.4 Comparison of intensity and value components.

$$H = \begin{cases} 0^\circ & \text{if } Max = Min \\ 60^\circ \times \frac{G - B}{Max - Min} + 0^\circ, & \text{if } Max = R \text{ and } G \geq B \\ 60^\circ \times \frac{G - B}{Max - Min} + 360^\circ, & \text{if } Max = R \text{ and } G < B \\ 60^\circ \times \frac{B - R}{Max - Min} + 120^\circ, & \text{if } Max = G \\ 60^\circ \times \frac{R - G}{Max - Min} + 240^\circ, & \text{if } Max = B \end{cases} \quad (3.1)$$

$$S = \begin{cases} 0, & \text{if } Max = 0 \\ \frac{Max - Min}{Max}, & \text{otherwise} \end{cases} \quad (3.2)$$

$$V = Max \quad (3.3)$$

In our preliminary experiments, intensity and value components of HSV were compared, and it was found that the value component was better than intensity in this application, so the histogram of value component within the top one-third of an image is calculated as a local color feature in the first stage. Fig. 3.4 shows images of Mt. Fuji in sunny

and night-time situations, each of which original image and those intensity and value component are shown. It shows that the intensity of both images is similar, whereas the value is distinguishable. Eq. (3.1) to (3.3) shows the calculations of H, S, and V components from RGB values. The range of R, G, and B component is  $[0, 1]$ . *Max* means the greatest value of R, G, and B. *Min* means the smallest value of R, G, and B. The range of H is  $[0, 360]$ , and those of S and V are  $[0, 1]$ , respectively.

Table 3.1 Precision and recall values of night-time situation experimented on different threshold of value component.

Number of Cluster	Measure	Threshold of Value component				
		72	76	80	84	88
2 clusters	Precision (%)	100.00	90.90	91.67	92.00	88.89
	Recall (%)	46.67	66.67	73.33	76.67	80.00

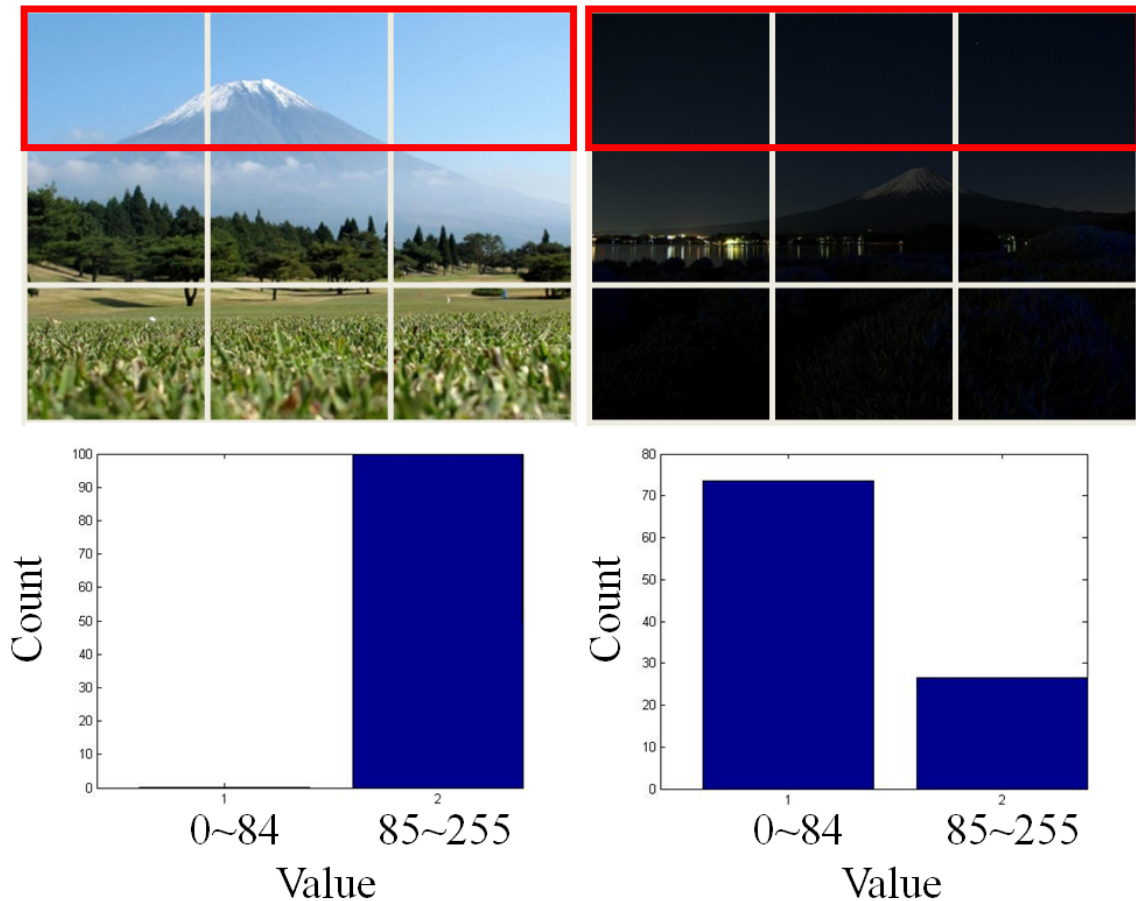


Fig. 3.5 Color histograms of value component calculated from the top 1/3 region of example images in day-time and night-time situations.

Fig. 3.5 shows two examples of color histogram of value component, which is extracted from the top one-third of images in daytime and night-time situations. After converting the range of value component from  $[0, 1]$  to  $[0, 255]$ , the threshold of value component is set to 84, which is determined based on results of preliminary experiments as shown in Table 3.1. After thresholding, a histogram with 2 bins is calculated as input to K-means clustering. Clusters are obtained by setting K at 2. The cluster with the higher value in a smaller bin is finally considered as a night-time situation, because the value component of night-time images in sky region is almost less than the threshold value. The illustration of situation discrimination in the first stage is shown in Fig. 3.6. The centroid of cluster 1 contains higher value in bin 1. Therefore, the cluster 1 (red part) is considered as night-time situation.

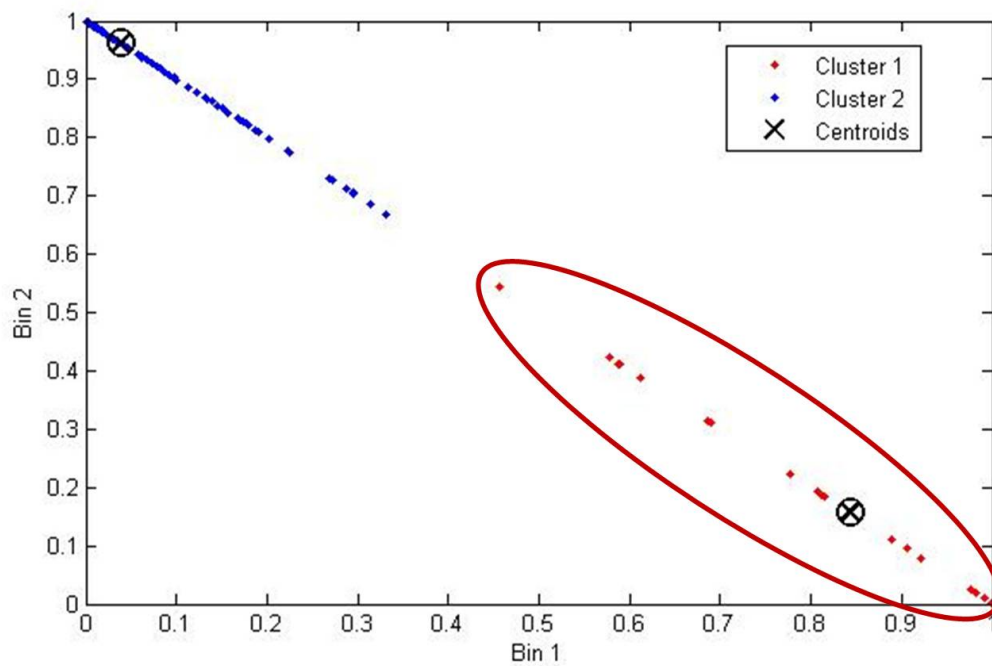


Fig. 3.6 Illustration of situation discrimination in the first stage.

### 3.4. Processing in the Second Stage

In the second stage, further grouping of other images, which are contained in the cluster other than the night-time one, into sunrise/sunset and daytime situations are considered. ROIs (regions of interest) in this stage include the sky region. The top one-third of an image usually contains other objects than the sky, however, such as clouds and mountain peaks, that affect color features of the region. Compared to the difference between night-time images and other images, the difference between sunrise/sunset images and daytime images is supposed to be small. In order to extract sky regions more exactly than what was done in 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 the block of local color feature extraction in Fig. 3.7:

- Step (1) Apply Canny edge detection [18] to obtain edge region ( $R_e$ ) of an input image.
- Step (2) Dilate  $R_e$  with a 5x5 kernel by using a morphology operation.
- Step (3) Reverse the dilated  $R_e$  to become non-edge region ( $R_{ne}$ ).
- Step (4) Get the global image threshold of a value component by using Otsu's method [19] and convert input images to binary images.
- Step (5) Get a binary image as the intersection of  $R_{ne}$  and the binary image from step (4).
- Step (6) Apply 8-connectivity to the binary image from step (5).
- Step (7) Get the binary image of the maximal region.
- Step (8) Obtain the ROI by dilating the binary image from step (7) with a 5x5 kernel.
- Step (9) Extract the color feature within the ROI.

Step (1) supposes that the land region of an image contains more complicated edges or texture than the sky region. Canny edge detection is therefore applied to find edges.

In order to achieve the optimal edge detection, Canny proposed a multi-stage algorithm. With the application of Gaussian filter for noise reduction, finding the intensity and gradient of images, non-maximum suppression for obtaining thin edges, and hysteresis thresholding with two thresholds, the multi-stage algorithm can detect a wide range of edges in images.



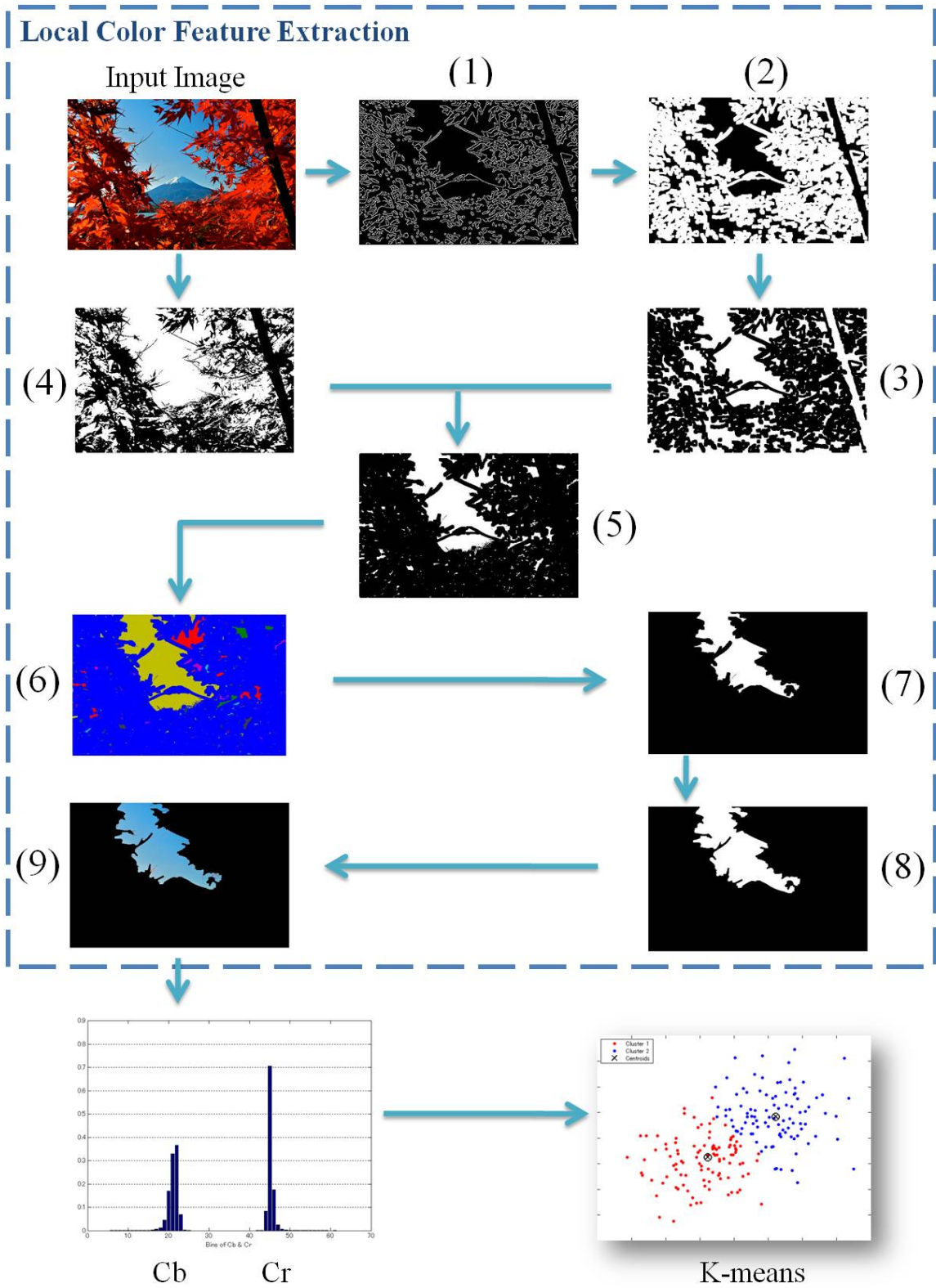


Fig. 3.7 Segmentation flow of ROI and example diagrams of histogram and K-means.



In Fig. 3.7 (1), the white region corresponds to detected edges. By applying steps (2) and (3), most objects such as land, buildings, and plants can be eliminated. Sometimes, however, edge detection is unavailable, for example, when the color of the land region is dark and this may lead to the extraction of an incorrect ROI. In order to avoid such problems, Otsu's method is applied in step (4) to find a global image threshold of a value component for eliminating dark land regions. The Otsu's method is generally used to find an adaptive threshold for reduction of a gray level image to a binary image. Step (5) intersects binary images obtained in steps (3) and (4), i.e., pixels that are white in both images are colored white in the resulting image.

In order to get individual connected regions, 8-connectivity is applied in step (6). Each connected region is labeled with a different value, which is indicated in different colors as shown in Fig. 3.7 (6). Step (7) extracts the maximal region that is the basis of the ROI to be extracted. In Fig. 3.7 (7), the white region is the maximal region. Dilation is performed on the binary image of the maximal region to get a more precise ROI in step (8). Finally, color features within the ROI finally are extracted for clustering.

$$\begin{bmatrix} Y \\ Cb \\ Cr \end{bmatrix} = \begin{bmatrix} 65.481 & 128.553 & 24.966 \\ -37.797 & -74.203 & 112.00 \\ 112.00 & -93.786 & -18.214 \end{bmatrix} \begin{bmatrix} R \\ G \\ B \end{bmatrix} + \begin{bmatrix} 16 \\ 128 \\ 128 \end{bmatrix} \quad (3.4)$$

It is observed that sunset images have mostly yellow and red colors in sky regions, while others tend to have blue and white colors in sky regions. Based on this observation, Cb and Cr components of YCbCr space within a ROI are extracted as local color features at this stage. Eq. (3.4) shows the conversion from RGB components to YCbCr components. The number of bins is set to 128 and clusters to 8, respectively, according to the best results in preliminary experiments. Table 3.2 and Table 3.3 show the result of preliminary experiments using RGB and CbCr components respectively. It is seen that CbCr components can obtain better result than RGB components for discriminating sunrise/sunset images from others.

Table 3.2 Precision and recall values of sunrise/sunset situation using RGB components.

Number of Cluster	RGB Components (Precision/Recall (%))				
	4 bins	8 bins	16 bins	32 bins	64 bins
4 clusters	25.64/34.48	27.08/44.83	19.7/44.83	21.21/48.28	20.00/44.83
8 clusters	100.00/31.03	95.45/72.41	100.00/58.62	92.00/78.31	100.00/10.34
16 clusters	100.00/24.14	100.00/24.14	100.00/24.14	100.00/27.59	100.00/44.83

Table 3.3 Precision and recall values of sunrise/sunset situation using CbCr components.

Number of Cluster	CbCr Components (Precision/Recall (%))				
	16 bins	32 bins	64 bins	128 bins	256 bins
4 clusters	96.00/82.76	88.00/75.86	23.91/75.86	23.66/75.86	25.00/79.31
8 clusters	92.31/41.38	95.65/75.86	93.33/96.55	93.55/96.67	92.00/79.31
16 clusters	92.31/41.38	100.00/58.62	92.31/41.38	100.00/65.52	100.00/75.86

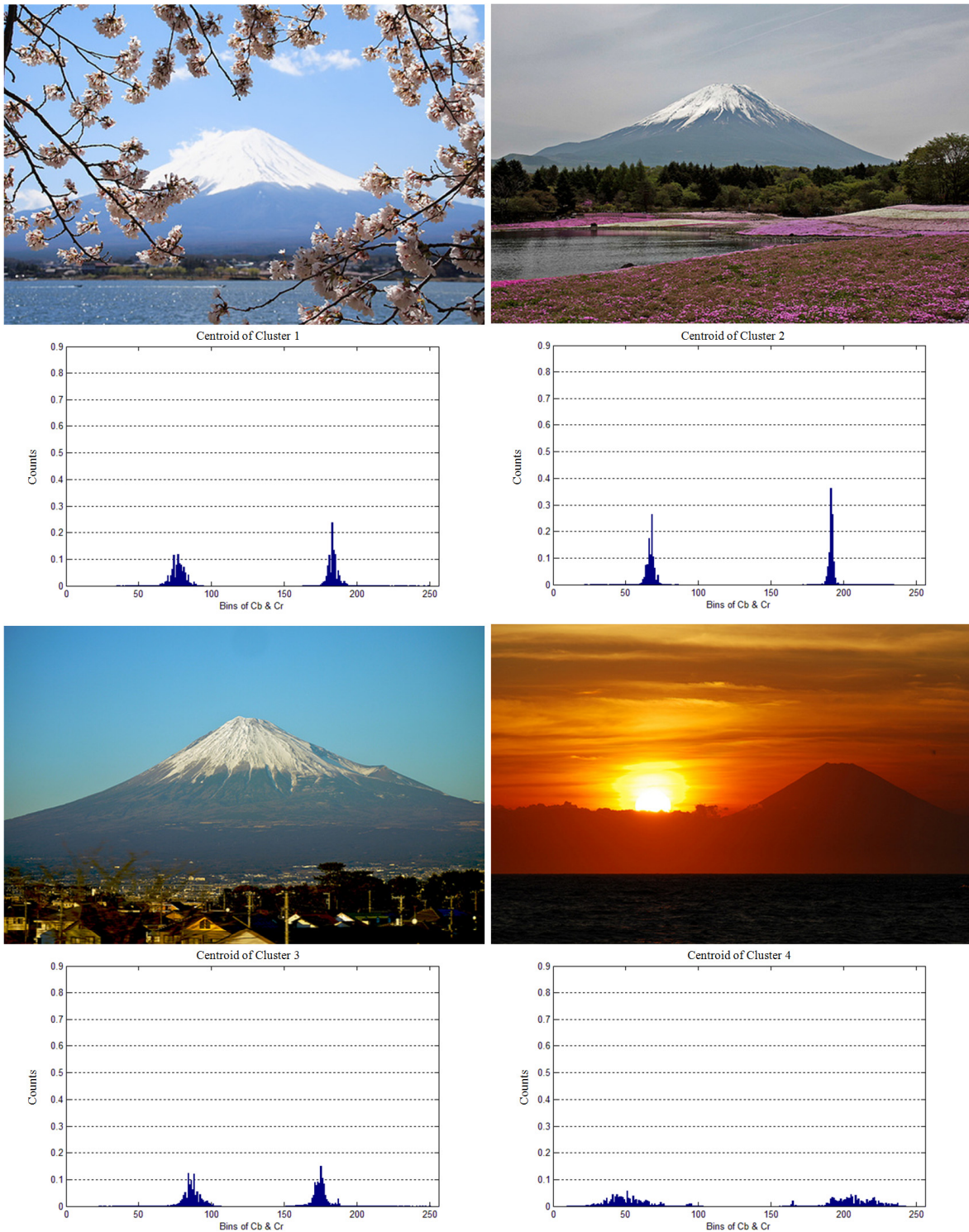


Fig. 3.8 Sample images in obtained cluster 1 to 4 and histograms of Cb & Cr values of their centroids in the second stage.

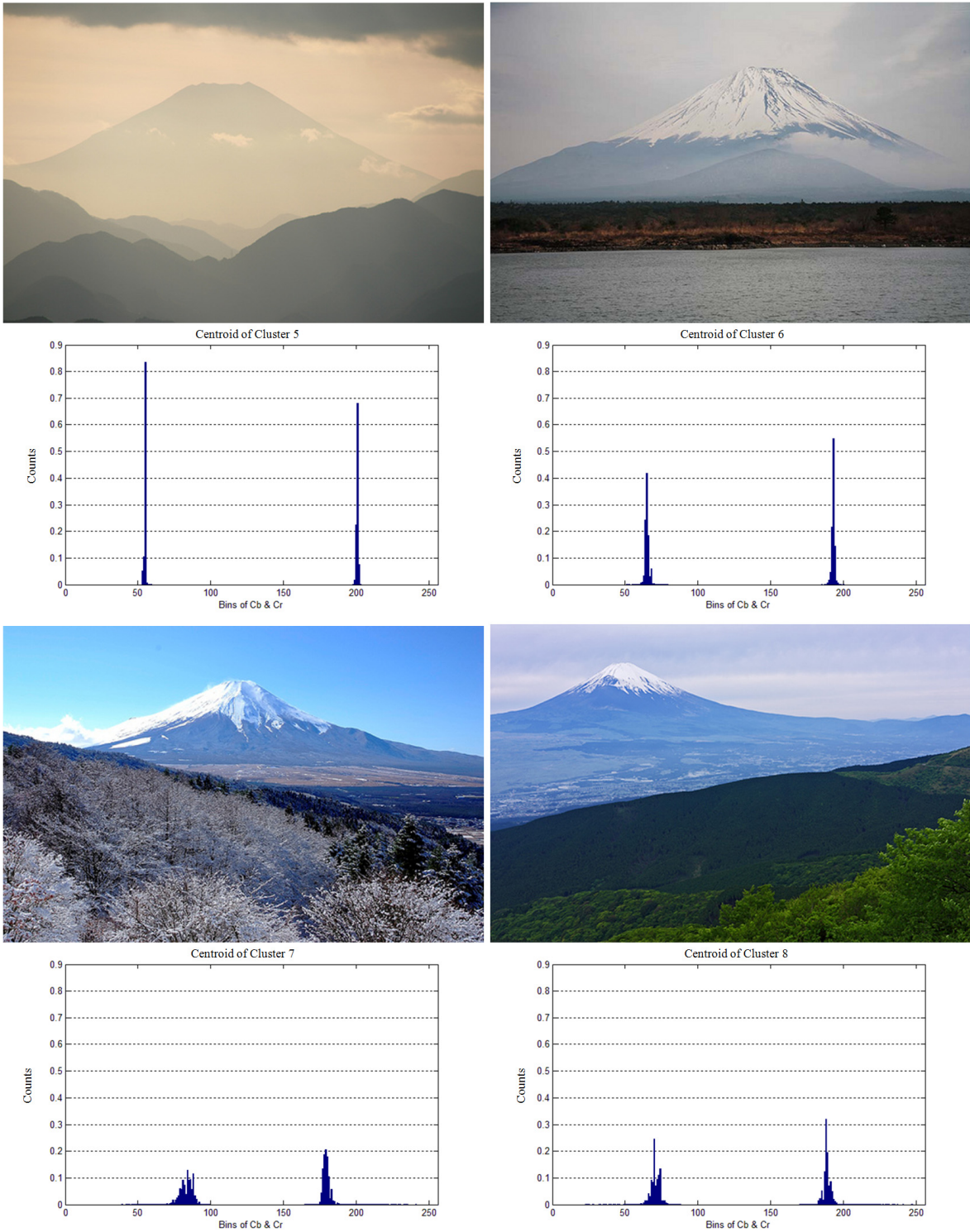


Fig. 3.9 Sample images in obtained cluster 5 to 8 and histograms of Cb & Cr values of their centroids in the second stage.

Fig. 3.8 and Fig. 3.9 show examples of 8 clusters including sunrise/sunset ones obtained. The figure consists of sample images in clusters and histograms of their centroids. Note that the sunrise/sunset cluster in Fig. 3.8 has smaller peaks than other clusters. The cluster having the smallest peak in the histogram of a centroid is therefore selected as a sunrise/sunset cluster.

### 3.5. Processing in the Third Stage

The purpose of the final stage is to group remaining images from the second stage into cloudy and sunshiny situations. The ROI is segmented by using the same method as that used in the second stage. Different combinations of HSV and YCbCr space were tested in preliminary experiments. Table 3.4 and Table 3.5 show the result of preliminary experiments using saturation and CbCr components respectively. It is seen that CbCr components can obtain better result than saturation component for discriminating cloudy images from sunshiny. Therefore, Cb and Cr components were also employed as features in this stage. The number of bins is set to 32 and clusters to 8, respectively, according to the best results in preliminary experiments.

Table 3.4 Precision and recall values of cloudy situation using saturation component.

Number of Cluster	Saturation Component (Precision/Recall (%))		
	32 bins	64 bins	128 bins
4 clusters	88.46/53.49	80.00/78.26	88.46/53.49
8 clusters	85.00/39.53	84.00/48.84	83.33/46.51

Table 3.5 Precision and recall values of cloudy situation using CbCr components.

Number of Cluster	CbCr Components (Precision/Recall (%))		
	32 bins	64 bins	128 bins
4 clusters	90.58/93.28	95.93/88.06	94.44/88.81
8 clusters	91.30/94.03	84.87/96.27	93.23/92.54

After K-means clustering, peak values of cluster centroids in the Cb component are compared to mean values of peak values for all cluster centroids. As shown in Fig. 3.8 and Fig. 3.9, it is observed that the distribution of Cb component in cloudy situation is narrower than that one in sunshiny situation. Therefore, the cluster whose centroid has higher peak values in the Cb component than the mean value is selected as a cluster for a cloudy situation. Multiple clusters satisfying this condition are merged into one cluster. The rest of the extracted clusters are also merged and considered to be sunshiny cluster.

### 3.6. Experiments

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

The proposed method is implemented using MATLAB. We collected images of 7 sightseeing spots on Flickr, from which images corresponding to a situation (night-time, sunrise/sunset, cloudy and sunshiny) were selected and labeled manually. The range of height and width in resolution is [300, 500]. In order to evaluate the effectiveness and limitations of the proposed method for various kinds of sightseeing spots, spots with different characteristics were selected, that is, Mt. Fuji and Mt. Takao were selected as typical natural scenes outside the city. Although Meiji Shrine is also famous as a natural scene, it is located inside the city and artifacts such as shrines and torii gate also exist in the area.

Tokyo Tower, Daiba, and Tokyo’s Rainbow Bridge were selected as typical spots that were artifacts. In terms of the distribution of images over situations, some spots contained very small numbers of images in certain situations, i.e., sunrise/sunset images at Sensoji and Meiji Shrine.

Five people took part in the labeling process, and an image was finally labeled only when all participants agreed on the label. Table 3.6 summarizes test dataset elements.

Table 3.6 Test dataset summary.

Search Words on Flickr	Number of Images				Total Labeled Images
	Night-time	Sunrise/Sunset	Cloudy	Sunshiny	
Mt. Fuji	30	30	46	134	240
Tokyo Tower	30	30	30	30	120
Daiba	118	69	57	154	398
Sensoji	93	17	226	217	553
Meiji Shrine	42	4	145	113	304
Mt. Takao	29	48	141	141	359
Rainbow Bridge Tokyo	149	94	50	106	399

In order to evaluate the performance of the proposed method, we applied the measures

of precision, recall, and F-measure commonly used in information retrieval. The precision (Eq. (3.5)) is measured by computing the ratio of the number of relevant images in a cluster divided by the total number of images in the cluster. The recall (Eq. (3.6)) 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. (3.7)) is a balanced mean between precision and recall.

$$precision = \frac{\#of\ relevant\ images\ in\ a\ cluster}{\#of\ images\ in\ a\ cluster} \quad (3.5)$$

$$recall = \frac{\#of\ relevant\ images\ in\ a\ cluster}{\#of\ relevant\ images} \quad (3.6)$$

$$F = 2 \times \frac{precision \times recall}{precision + recall} \quad (3.7)$$

Table 3.7 shows results for discriminating night-time images in the first stage for Mt. Fuji and Tokyo Tower. Values of precision and recall are measured by the proposed method (the value component of HSV space within the top one-third region of image), local intensity (intensity within the top one-third region of image), global value (the value component of HSV space within the whole image) and global intensity (intensity within the whole image). Note that the value component performs better than intensity. It is observed that the performance of intensity gets worse when a sunshiny image contains deep blue or dark sky as shown in Fig. 3.4.

Table 3.7 Precision and recall values of night-time situation in the first stage.

Method	Mt. Fuji		Tokyo Tower	
	Precision (%)	Recall (%)	Precision (%)	Recall (%)
Proposed	92.00	76.67	96.67	96.67
Local Intensity	87.50	70.00	96.43	90.00
Global Value	77.42	80.00	78.38	96.67
Global Intensity	65.00	86.67	79.41	90.00



Note also that results of local color feature are better than for global color features. In results for Mt. Fuji, although recall of the proposed method is lower than for methods using the global color feature, the precision of proposed method is much higher than for these methods. In the case of Mt. Fuji, it is observed that some images were taken at night using high exposure compensation. The sky areas of such images tended to be brighter than usual, which leads to incorrect classification and decreases recall values. The number of images for a sightseeing spot obtained from the Web is usually huge, so we consider precision more important than recall.

In order to show the performance of the second and the third stages, the proposed method was compared to the following methods:

- w/o-Otsu: This method is the same as the proposed method but skips step (4), as shown in Fig. 3.7.
- Global-CbCr: This method uses Cb and Cr components within global images.
- ROI-RGB: This method uses R, G and B components within the same ROI as that for the proposed method.

Table 3.8 Precision and recall values of sunrise/sunset situations in the second stage.

Method	Mt. Fuji		Tokyo Tower	
	Precision (%)	Recall (%)	Precision (%)	Recall (%)
Proposed	93.55	96.67	96.67	100.00
w/o-Otsu	92.86	86.67	96.55	93.33
Global-CbCr	92.86	43.33	95.00	63.33
ROI-RGB	90.91	66.67	75.00	70.00

Table 3.9 Precision and recall values of cloudy situations in the third stage.

Method	Mt. Fuji		Tokyo Tower	
	Precision (%)	Recall (%)	Precision (%)	Recall (%)
Proposed	84.78	84.78	84.85	93.33
w/o-Otsu	81.82	78.26	82.35	93.33
Global-CbCr	82.05	69.57	85.19	76.67
ROI-RGB	76.92	43.48	75.00	10.00

Table 3.10 Precision and recall values of sunshiny situations in the third stage.

Method	Mt. Fuji		Tokyo Tower	
	Precision (%)	Recall (%)	Precision (%)	Recall (%)
Proposed	91.30	94.03	92.31	80.00
w/o-Otsu	90.00	94.03	92.00	76.67
Global-CbCr	86.90	94.03	81.25	86.67
ROI-RGB	79.11	93.28	52.73	96.67

Table 3.8 shows results for sunrise/sunset situations in the second stage. Table 3.9 and Table 3.10 show results for cloudy and sunshiny situations respectively in the third stage. These tables show results for Mt. Fuji and Tokyo Tower. The proposed method gets the best results both for precision and recall in these three situations, except in the case of cloudy and sunshiny situations for Tokyo Tower. The precision of Global-CbCr is thus a little higher than for the proposed method in Table 3.9. Although precision is more important as mentioned above, the amount of difference in precision in this case is only 0.34 percentage points but recall for the proposed method is about 16.7 percentage points higher than for Global-CbCr. The proposed method therefore still performs better than Global-CbCr.

Results for w/o-Otsu indicate that adding thresholding of a global image with Otsu's method is effective. A comparison of the proposed method and Global-CbCr shows that feature extraction from the ROI is effective. A comparison of the proposed method and ROI-RGB shows that the Cb and Cr component is more suitable than RGB space in the clustering of sunrise/sunset, cloudy and sunshiny situations. When we compare results for the proposed method for different situations, results of cloudy situations are worse than for other situations because the color feature extracted from cloudy images contains mostly white and gray, whereas Cb and Cr represent blue-difference and red-difference chroma components, respectively.

Because the proposed method uses K-means clustering in each stage, results would be different in each execution, so experiments were conducted to calculate the best value, average and standard deviation (SD) by running 10 times. Table 3.11 shows results, which show that standard deviation is relatively small compared to average values in most cases.

Table 3.11 Best value, average, and standard deviation for precision and recall.

Spot	Statistics	Night-time (Precision /Recall (%))	Sunrise/Sunset (Precision /Recall (%))	Cloudy (Precision /Recall (%))	Sunshiny (Precision /Recall (%))
Mt. Fuji	Best	92.00/76.67	93.55/96.67	84.78/86.96	92.17/94.03
	Average	92.00/76.67	93.00/93.00	67.23/85.22	91.37/84.63
	S.D.	0.00/0.00	0.86/4.58	11.71/1.31	0.55/6.33
Tokyo Tower	Best	96.67/96.67	96.77/100.00	84.85/93.33	92.31/80.00
	Average	96.67/96.67	90.33/98.34	84.85/93.33	92.31/80.00
	S.D.	0.00/0.00	7.92/1.67	0.00/0.00	0.00/0.00
Daiba	Best	94.83/93.22	90.74/72.46	88.10/87.72	86.67/95.45
	Average	94.83/93.22	89.52/68.84	84.15/67.19	79.69/93.76
	S.D.	0.00/0.00	1.69/3.05	8.13/6.84	2.35/3.57
Sensoji	Best	88.30/89.25	12.66/58.82	94.34/88.50	86.31/66.82
	Average	88.30/89.25	12.57/58.82	94.34/88.50	86.31/66.82
	S.D.	0.00/0.00	0.18/0.00	0.00/0.00	0.00/0.00
Meiji Shrine	Best	75.47/95.24	6.67/50.00	90.08/91.03	86.57/58.41
	Average	75.47/95.24	6.67/50.00	87.72/86.14	79.22/54.51
	S.D.	0.00/0.00	0.00/0.00	2.02/4.01	5.44/3.29
Mt. Takao	Best	67.50/93.10	80.43/77.08	89.93/88.65	89.55/85.11
	Average	67.50/93.10	80.43/77.08	89.37/88.15	89.02/84.54
	S.D.	0.00/0.00	0.00/0.00	1.69/1.49	1.60/1.70
Rainbow Bridge Tokyo	Best	95.27/94.63	100.00/72.34	71.64/98.00	86.49/90.57
	Average	95.27/94.63	99.28/69.15	68.99/97.80	86.15/90.29
	S.D.	0.00/0.00	0.73/3.19	0.98/0.60	0.89/0.44

In order to verify the effectiveness of the proposed hierarchical classification, it is compared to ordinary K-means using the same color features. That is, each image is represented with 258 attributes including values, Cb and Cr components. As noted in Section 3, the proposed method selects one cluster as a night-time situation and one cluster as a sunrise/sunset situation in the first and second stages respectively. In the third stage, 8 clusters are separated into cloudy and sunshiny situations, so the number of clusters is set to 10. After clustering, the same criterion of discrimination for each situation as for the proposed method is applied.

We call this classification method the baseline method hereafter. Table 3.12 shows results of comparison in average precision and recall for 10 runs.

The comparison of average F-measures is also shown in Table 3.13, where the highest score is marked with an asterisk (\*). Note that the proposed method gets better results than the baseline method for all 4 situations for Daiba, Mt. Takao, and Tokyo's Rainbow Bridge.

The proposed method also gets better results than the baseline method for 3 situations in four other spots.

Table 3.12 Comparison of proposed method with baseline method in average of precision and recall.

Spot	Method	Night-time (Precision /Recall (%))	Sunrise/Sunset (Precision /Recall (%))	Cloudy (Precision /Recall (%))	Sunshiny (Precision /Recall (%))
Mt. Fuji	Proposed	92.00/76.67	93.00/93.00	67.23/85.22	91.37/84.63
	Baseline	98.73/38.00	91.99/93.67	72.69/92.61	84.83/87.39
Tokyo Tower	Proposed	96.67/96.67	90.33/98.34	84.85/93.33	92.31/80.00
	Baseline	100.00/35.33	93.32/99.33	83.79/77.00	57.69/88.34
Daiba	Proposed	94.83/93.22	89.52/68.84	84.15/67.19	79.69/93.76
	Baseline	96.21/23.64	76.06/52.03	66.65/83.34	54.42/88.18
Sensoji	Proposed	88.30/89.25	12.57/58.82	94.34/88.50	86.31/66.82
	Baseline	96.41/86.67	12.91/51.76	91.81/75.13	70.52/70.14
Meiji Shrine	Proposed	75.47/95.24	6.67/50.00	87.72/86.14	79.22/54.51
	Baseline	81.82/21.43	2.69/25.00	88.23/68.28	65.28/82.83
Mt. Takao	Proposed	67.50/93.10	80.43/77.08	89.37/88.15	89.02/84.54
	Baseline	78.24/60.00	76.10/77.08	88.13/64.68	69.00/88.58
Rainbow Bridge Tokyo	Proposed	95.27/94.63	99.28/69.15	68.99/97.80	86.15/90.29
	Baseline	100.00/21.28	4.38/3.51	68.43/66.40	41.83/92.83

Table 3.13 Comparison of proposed method with baseline method in average of F-measure.

Spot	Method	Night-time (F-measure (%))	Sunrise/Sunset (F-measure (%))	Cloudy (F-measure (%))	Sunshiny (F-measure (%))
Mt. Fuji	Proposed	83.64 *	93.00 *	75.16	87.87 *
	Baseline	54.88	92.82	81.45 *	86.09
Tokyo Tower	Proposed	96.67 *	94.16	88.89 *	85.72 *
	Baseline	52.21	96.23 *	80.25	69.80
Daiba	Proposed	94.02 *	77.83 *	74.72 *	86.15 *
	Baseline	37.95	61.79	74.07	67.30
Sensoji	Proposed	88.77	20.71 *	91.33 *	75.32 *
	Baseline	91.28 *	20.67	82.64	70.33
Meiji Shrine	Proposed	84.21 *	11.77 *	86.92 *	64.58
	Baseline	33.96	4.86	76.98	73.02 *
Mt. Takao	Proposed	78.26 *	78.72 *	88.76 *	86.72 *
	Baseline	67.92	76.59	74.61	77.57
Rainbow Bridge Tokyo	Proposed	94.95 *	81.52 *	80.91 *	88.17 *
	Baseline	35.09	3.90	67.40	57.67

Note that the baseline method gets much lower recall than the proposed method for night-time situations in most cases. This is because it generates 10 clusters and the size of each cluster tends to be small. The proposed method, in contrast, generates only 2 clusters at the first stage for night-time situations, so it successfully groups night-time images in one cluster. Note also that the baseline method gets lower precision than the proposed method in sunshiny situations, and sometimes obtains very low performance such as in sunrise/sunset situations for Tokyo's Rainbow Bridge. Such low performance is mainly caused by night-time images being grouped into the same clusters as sunshiny images (and also as sunrise/sunshiny images in the case of Tokyo's Rainbow Bridge). The proposed method avoid getting such low performance, however, because of hierarchical classification, because night-time images are caught in the first stage and do not affect clusters of other situations.

$$\mu = \frac{1}{N} \sum_{i=1}^N x_i \quad (3.8)$$

$$\sigma^2 = \frac{\sum_{i=1}^N (x_i - \mu)^2}{N} \quad (3.9)$$

In order to measure the computational costs, the processing time of proposed method is measured for all situations. The arithmetic mean  $\mu$  (Eq. (3.8)) and variance  $\sigma^2$  (Eq. (3.9)) is calculated, where  $N$  means the number of procedure running and  $x_i$  means the computational costs of the  $i^{\text{th}}$  procedure running. Table 3.14 shows result of average computational costs, i.e. mean and variance for 5 runs and total number of images in each spot. The processing time depends on the number of images and the result shows that proposed method can process about 1.5 to 2 images in 1 second. It is also shown the variation of computational costs is under acceptable range. Since the major purpose of this application is to classify images into different situations in off-line processing, the real-time speed processing is not needed. Therefore, it is acceptable to classify 1000 images in 10 minutes.

Table 3.14 Total number of images and computational costs of proposed method in average and variance.

Spot	Average (Second)	Variance	Total Number of Images
Mt. Fuji	163.96	1.44	240
Tokyo Tower	59.29	1.25	120
Daiba	201.78	6.86	398
Sensoji	345.36	3.52	553
Meiji Shrine	191.18	8.25	304
Mt. Takao	240.63	5.22	359
Rainbow Bridge Tokyo	189.10	11.47	399

Fig. 3.10 shows sample images of Mt. Fuji contained in clusters for each situation that are obtained by the proposed method. Note that outdoor sightseeing images in night-time (a) and other (b) situations are easy to discriminate among by using the brightness of the top one-third of the region. After night-time images are isolated, it becomes difficult to discriminate sunrise/sunset (c) from daytime (d) because images may contain green or red leaves in the top one-third of the region. The specific ROI and available color space are therefore proposed. By extracting ROI, we get much more easily distinguishable colors such as orange in sunrise/sunset situations. The proposed method finally discriminates sunrise/sunset images (c) from daytime images (d) and cloudy images (e) from sunshiny images (f) in the second stage and the third stages, respectively.

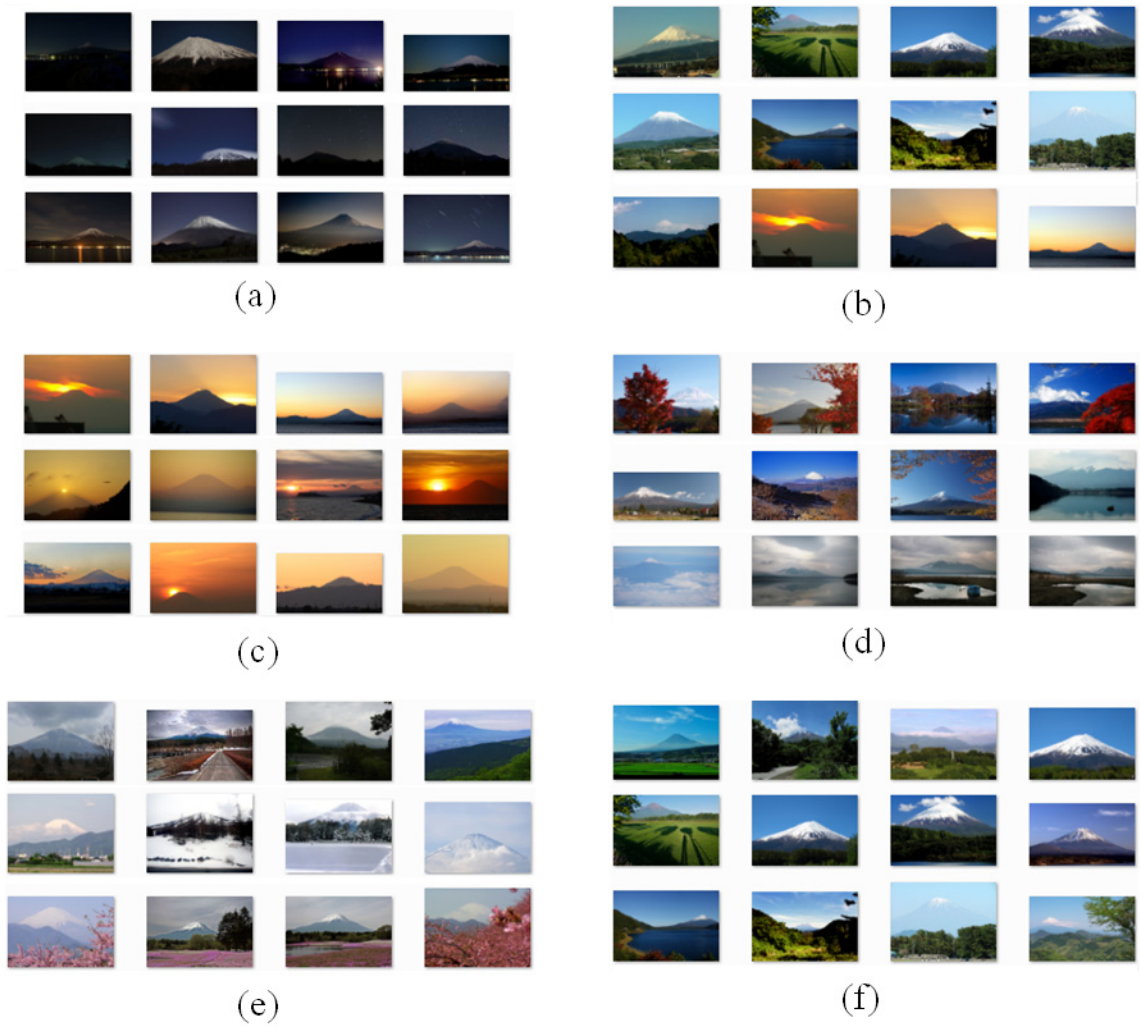


Fig. 3.10 Images of individual situations for Mt. Fuji. (a) night-time and (b) other situations in the first stage. (c) sunrise/sunset situation and (d) daytime in the second stage. (e) cloudy and (f) sunshiny situations in the third stage.

## **4. Hybrid Approach Based on Visual and Metadata Features for Image Classification Targeting Time-Related Situations**

### **4.1. Utilization of Tag Information for Time-Related Situations**

This chapter proposes image classification method targeting time-related situation which includes night-time, sunrise/sunset, daytime (cloudy and sunshiny).

The image classification method proposed in Chapter 3 is based on the visual perspective of situations. It is supposed that people recognize images belong to a certain situation according to the changes of brightness or color in local region. Therefore, color features extracted from ROI, i.e. the sky region is expected to be useful for classifying weather-related and time-related situations. However, some limitations exist in content-based classification method. Some images were mis-classified due to the sky region being covered by some objects such as roofs and trees. The light reflection also influences the result of classification. However, in the case of time-related situation, different situations such as night, sunrise/sunset, and daytime (cloudy and sunshiny) corresponds to different time of the 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 chapter proposes a hybrid approach for improving the accuracy of classifying time-related 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.



## 4.2. Overall Procedure

The overall processing flow as shown in Fig. 4.1 consists of three clustering and three filtering processes. It is observed in the collected image dataset that the classification of different situations by time information only has some challenges. For example, 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 proposed in Chapter 3 is applied to divide the collected images into night-time cluster and other images at the beginning. By using tag information, the images in night-time cluster are verified with the night-time filter, which separates mis-clustered images from night-time images.

Other images which are not classified in night-time cluster and the mis-clustered images will be considered as input for next clustering process. This round of process corresponds to the first stage as shown in Fig. 4.1. The other images are further divided into sunrise/sunset and daytime clusters at second stage by applying content-based image classification, which is followed by time filter for sunrise/sunset. At the 3rd stage, the cloudy images are discriminated from sunshiny images by content-based classification and daytime filter for verification.

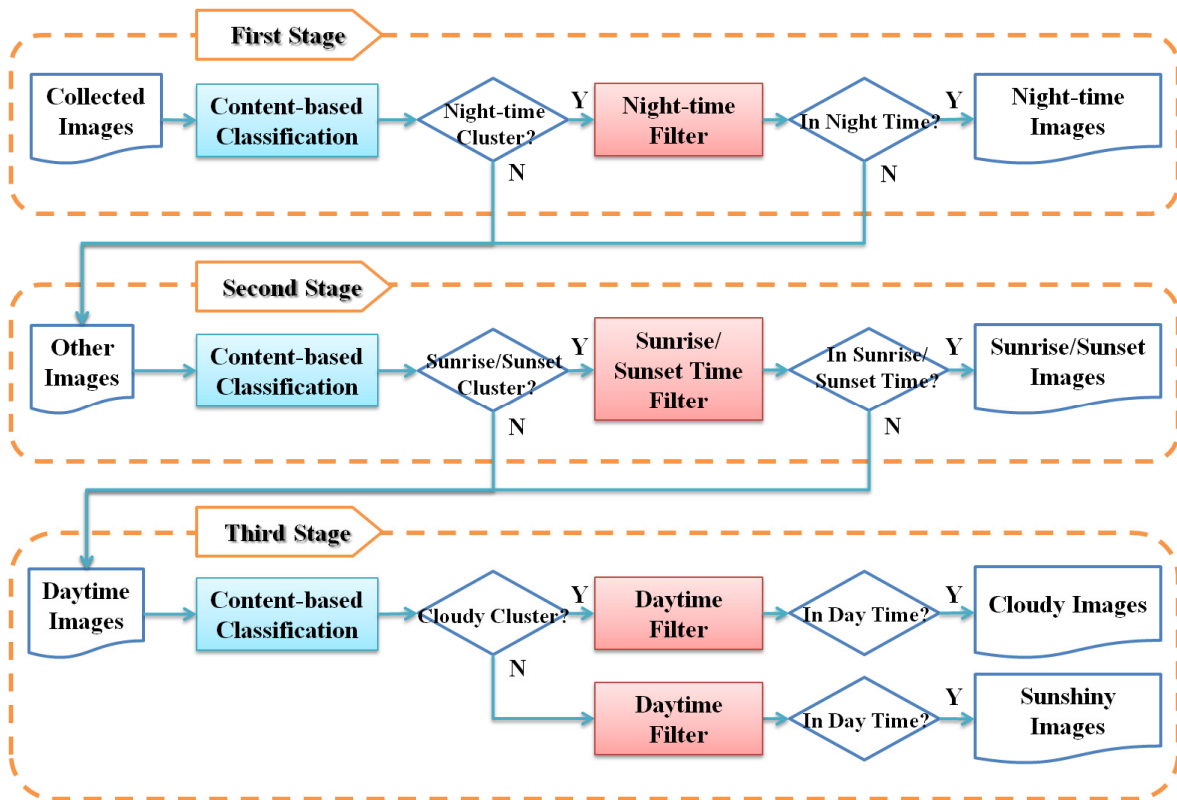


Fig. 4.1 Overall processing flow.

### 4.3. Definition of Time Window

The situation handled in this chapter is time-related one which includes night, sunrise/sunset, and daytime (cloudy and sunshiny) as mentioned above. As different situations correspond to different time of the day, except that cloudy and sunshiny situation, this chapter 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 chapter) from the website of National Astronomical Observatory of Japan<sup>5</sup>. Table 4.1 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 to Mt. Fuji. In the case of Arashiyama, it refers to the times of Kyoto.

Table 4.1 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

<sup>5</sup> [http://eco.mtk.nao.ac.jp/cgi-bin/koyomi/koyomix\\_en.cgi](http://eco.mtk.nao.ac.jp/cgi-bin/koyomi/koyomix_en.cgi)

Table 4.2 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 & Sunshiny)	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-time	1:00 ~ 7:00,	1:00 ~ 8:00,	1:00 ~ 8:00,
	16:00 ~ 24:00	16:00 ~ 24:00	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 4.2. 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.

## 4.4. Hierarchical Classification with Time Filter

### 4.4.1. Processing at the First 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 [15]. 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 Chapter 3.



Fig. 4.2 Example of cloudy images that are mis-classified by content-based method but correctly classified by hybrid method.

After clustering and discrimination process, the night-time filter considering the time window of night situation is applied to verify night-time cluster's images. As noted in Sec. 4.3, 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. 4.2 shows an example of images that are mis-classified by content-based method but correctly classified by hybrid method. Two cloudy images are mis-classified into night-time situation because the top one-third region is covered by some objects such as roof and tree.

#### 4.4.2. Processing at the Second 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, which is described in Sec. 3.4. The method consists of the following 5 steps.

- Step (1) Apply Canny edge detection [18] to obtain edge region ( $R_e$ ) of an input image and then reverse the dilated  $R_e$  which is dilated with 5x5 kernel by morphology operation to obtain non-edge region ( $R_{ne}$ ).
- Step (2) Get a global image threshold of value component by Otsu's method [19] and convert input image to binary image.
- Step (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.
- Step (4) Get the binary image of maximal region and then obtain ROI by dilating the binary image with 5x5 kernel.
- Step (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.

Fig. 4.3 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 in terms of Cb and Cr components. 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.

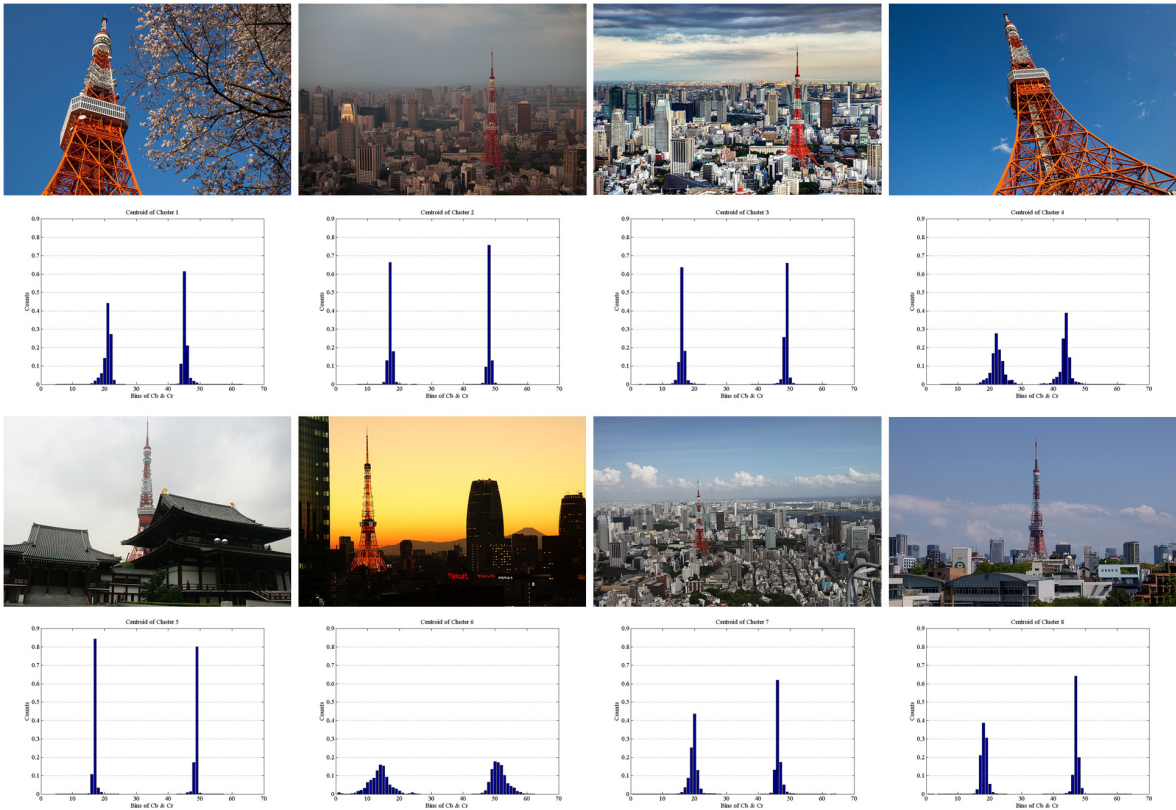


Fig. 4.3 Sample images in obtained clusters and histogram of their centroids at second stage.





Fig. 4.4 Example of night-time images that are mis-classified by content-based method but correctly filtered and removed by hybrid method.

After clustering and discrimination process, the time filter considering the time window of sunrise/sunset situation such as shown in Table 4.2 is applied to verify the taken time of images in sunrise/sunset cluster to improve the accuracy in this stage. Fig. 4.4 shows an example of images that are mis-classified by content-based method but correctly filtered and removed by hybrid method. Because of light affection, two night-time images are mis-classified into sunrise/sunset situation. However, these kind of images can be filtered out by proposed time windows.

#### 4.4.3. Processing at the Third Stage

The purpose of the final stage is to group remaining images from the second stage into cloudy and sunny situations. The ROI is segmented by the same method as the second 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 sunny cluster.

After clustering and discrimination process, the time filter considering the time window of daytime situation such as shown in Table 4.2 is applied to verify the taken time of images in both cloudy and sunny clusters to improve the accuracy in this stage.

## 4.5. 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 geotag information, 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 4.3, 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 ( $35.658610 \pm 0.2$ ) and longitude ( $139.745447 \pm 0.2$ ) and then filter boundary of latitude ( $35.658610 \pm 0.00005$ ) and longitude ( $139.745447 \pm 0.00005$ ) was applied to eliminate the images which were taken inside the tower. Fig. 4.5 shows an example of collected and filtered region for Mt. Fuji.

Table 4.3 Parameter settings of image collection.

Query Text	Area	Central Point (Latitude, Longitude)	Collection Range	Filter Range	Total Images
Tokyo Tower	Tokyo	35.658610, 139.745447	0.2	0.00005	3,792
Mt. Fuji	Shizuoka	35.362596, 138.731232	0.4	0.07	2,409
Daiba	Tokyo	35.630599, 139.778449	0.05	0.0003	1,297
Sensoji	Tokyo	35.714751, 139.796685	0.05	0.0002	2,050
Meiji Shrine	Tokyo	35.676324, 139.69938	0.05	0.0003	1,981
Rainbow Bridge Tokyo	Tokyo	35.635604, 139.76635	0.2	0.00005	608
Arashiyama	Kyoto	35.015194, 135.677706	0.2	0	3700

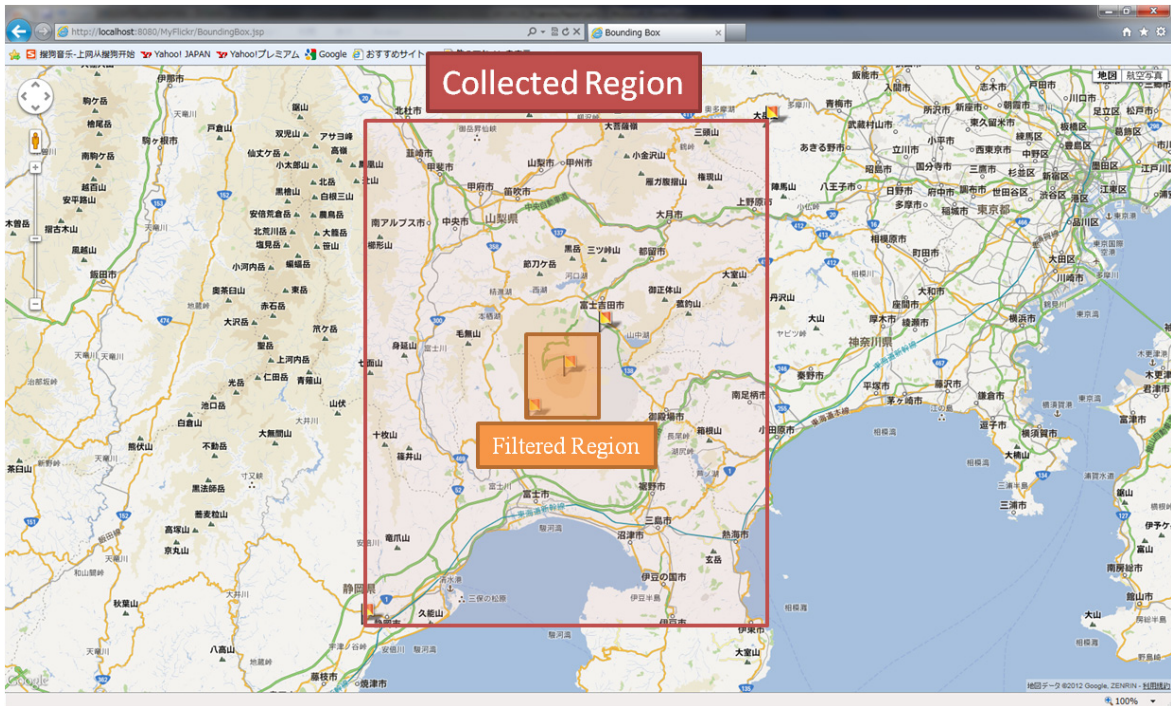


Fig. 4.5 Collected and filtered region of Mt. Fuji.

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-time, sunrise/sunset, cloudy and sunshiny) 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 4.4 summarizes the obtained test dataset.

In order to store sightseeing images and their metadata, a web application on a tomcat server is constructed to retrieve images and tag information from Flickr as shown in Fig. 4.6. Tags are stored on Fuseki server in RDF (Resource Description Framework) format according to the designed data model. The Fuseki<sup>6</sup> is a kind of SPARQL Endpoint provided by the Apache Jena project, which can store and retrieve data in RDF format. This application is also a prototype using the proposed method, i.e., the images in different situations can be displayed on map-based interface.

<sup>6</sup> [http://jena.apache.org/documentation/serving\\_data/index.html](http://jena.apache.org/documentation/serving_data/index.html)

Table 4.4 Test dataset.

Spot	Number of Images				Total Labeled Images
	Night-time	Sunrise/ Sunset	Cloudy	Sunshiny	
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

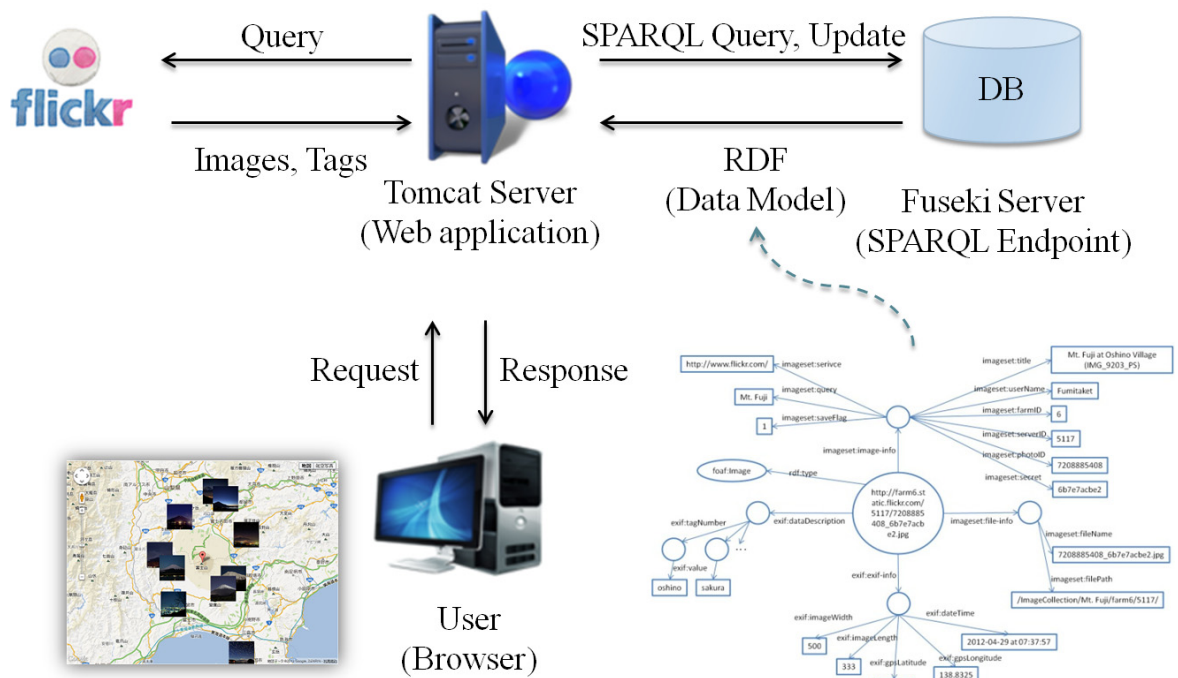


Fig. 4.6 System architecture for sightseeing images classification and rendering in different situations.

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 is measured by computing the ratio of number of relevant images in a cluster divided by the total num-

ber of images in the cluster. The recall is computed by dividing the number of relevant images in a cluster by the total number of relevant images in the dataset.

Table 4.5 Average precision and recall values (%) measured by different methods in each situation.

Spot	Method	Night-time (Precision/Recall)	Sunrise /Sunset (Precision/Recall)	Cloudy (Precision/Recall)	Sunshiny (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

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 as shown in Fig. 4.1.
- **Hybrid (proposed method):** this method performs content-based image classification first and then utilizes time windows to filter out outliers as shown in Fig. 4.1.

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 4.5 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-time and sunshiny 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.

In cloudy situation of Daiba and sunshiny 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 4.6 shows F-measure of both methods. 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.

Table 4.6 Comparison of hybrid method and content-based method in F-measure (%).

Spot	Method	Night-time	Sunrise/Sunset	Cloudy	Sunshiny
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	83.50
	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 *	88.50	82.75

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 4.4 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. Fig. 4.7 shows the sample images (top row) that are mis-classified into sunrise/sunset cluster and their ROI images (bottom row). The first and second sample images are labeled as cloudy and night respectively. It is seen that the color of cloud and light is similar to the sunrise/sunset color in the 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. 4.7 Sample images (top row) that are mis-classified into sunrise/sunset cluster and their ROI (bottom row). The first and second image is labeled as cloudy and night situation respectively.

Fig. 4.8 shows the classification results of Tokyo Tower and Mt. Fuji obtained by proposed method. It is observed the images are correctly classified into night-time (a), sunrise/sunset (b), cloudy (c), and sunshiny (d) situations.





Fig. 4.8 Sample images of (a) night-time, (b) sunset/sunrise, (c) cloudy, and (d) sunny situations obtained by proposed hybrid method ( (1) Tokyo Tower and (2) Mt. Fuji).

## 5. Identification of Season-Dependent Sightseeing Spots Based on Metadata-Derived Features and Image Processing

### 5.1. Identification of Season-Dependent Sightseeing Spots

This chapter focuses on season-related situations. As noted in Chapter 1, main characteristic of season-related situations is that scenery change by season-related ones will depend on a sightseeing spot. Even though two sightseeing spots are geographically close, it often happens one maybe season-dependent and the other not. For example, Fig. 5.1 shows several photos of Shinjuku Gyoen and Rainbow Bridge Tokyo which were taken in April, July, November, and January. It is obvious that the color of objects such as flowers and leaves in the photos of Shinjuku Gyoen varies with different seasons but objects in the photos of the Rainbow Bridge Tokyo do not have such kinds of variation. Although both of these two spots are located in Tokyo, the Shinjuku Gyoen and the Rainbow Bridge Tokyo are season-dependent and season-independent, respectively.

Spot	April	July	November	January
Shinjuku Gyoen				
Rainbow Bridge Tokyo				

Fig. 5.1 Example photos of Shinjuku Gyoen and Rainbow Bridge Tokyo taken in April, July, November, and January.

This thesis defines season-dependent spots as those of which scenery changes according to season. On the other hand, spots of which scenery does not change according to season are called season-independent spots. As shown in above-mentioned example, it is difficult to recognize a sightseeing spot as season dependent / independent just according to location only. Although the supervised learning with visual features is expected to work well for such a classification, it consumes a lot of time to download photos of various seasons and to apply image processing to many photos for each sightseeing spot. Therefore, this chapter proposes two-stage approach that employs a set of statistical data about a

sightseeing spot at the first stage and image processing at the second stage. Fig. 5.2 shows the overall processing flow of the proposed method.

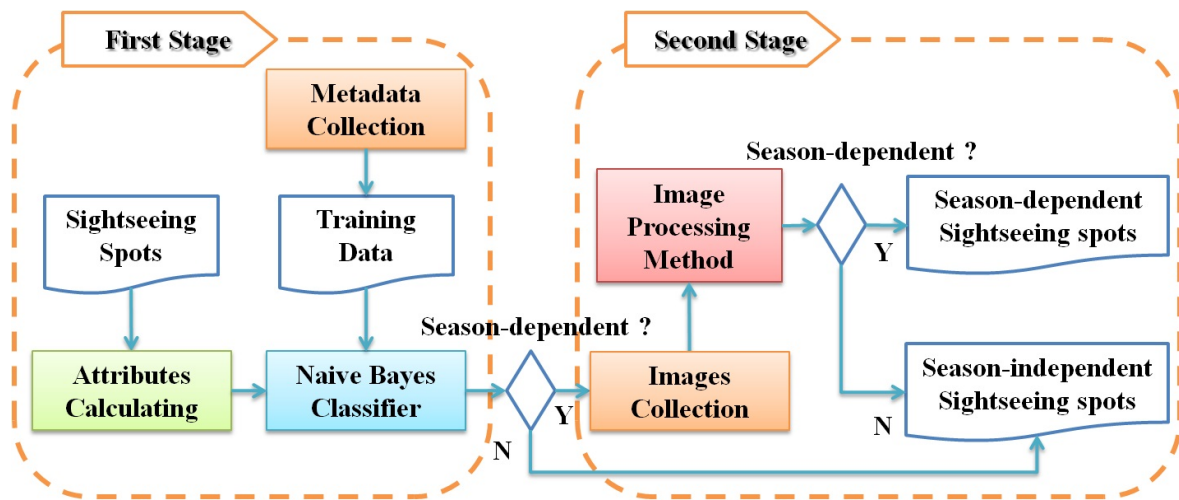


Fig. 5.2 Overall processing flow.

## 5.2. Machine Learning in the First Phase

A web album such as Flickr does not only provide sightseeing photos but also the information about statistics such as the number of photos and tourists. The tourists mean those who uploaded their photos taken at a sightseeing spot to the web album. Because the scene of a season at a sightseeing spot attracts many people and lets them take photos, it makes this kind of data meaningful. For example, one of season-dependent spots in Tokyo is Shinjuku Gyoen and one of season-independent ones is Rainbow Bridge Tokyo. Fig. 5.3 shows the monthly number of tourists in 2011 at Shinjuku Gyoen and Rainbow Bridge Tokyo, respectively. It is noted that Shinjuku Gyoen's tourists in April are much more than other month because it is famous for cherry blossoms in spring. On the other hand, the number of tourists in Rainbow Bridge Tokyo is not so large in most months.

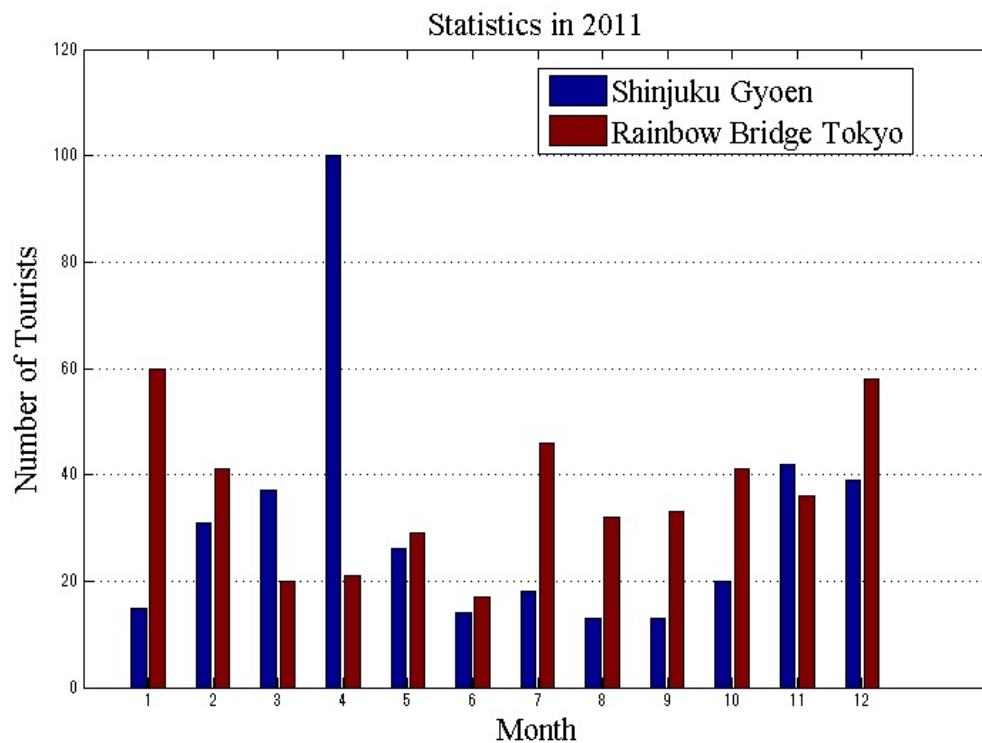


Fig. 5.3 Monthly number of tourists of Shinjuku Gyoen and Rainbow Bridge Tokyo in 2011.

Based on these considerations, this chapter employs the following ten attributes based on statistics for classifying sightseeing spots into season-dependent / independent ones. It is noted that the following attributes are based on statistics of recent 3 years.

- (A1) Average difference in normalized number of tourists between different years. At first 3 values are calculated, i.e., difference between 2009-2010 (A11), 2010-2011 (A12), and 2011-2009 (A13). The average of these 3 values is considered as the first attribute.
- (A2) Difference in the normalized average number of tourists between the highest month and the lowest month. An average number of tourists is calculated based on tourists from 2009 to 2011.
- (A3) Peak month that has the most tourists during 2009 to 2011. The range of this attribute is  $\{1,2,\dots,12\}$ .
- (A4) Month that has the least tourists during 2009 to 2011. The range of this attribute is  $\{1,2,\dots,12\}$ .
- (A5) The normalized maximum number of tourists during 2009 to 2011.
- (A6) The distance (in month) between the peak month and the second peak month.
- (A7) Total number of tourists in 2009 (A71), 2010 (A72), and 2011 (A73).
- (A8) Average difference in normalized number of tourists during 2009 to 2011 between a target spot and average of all spots.

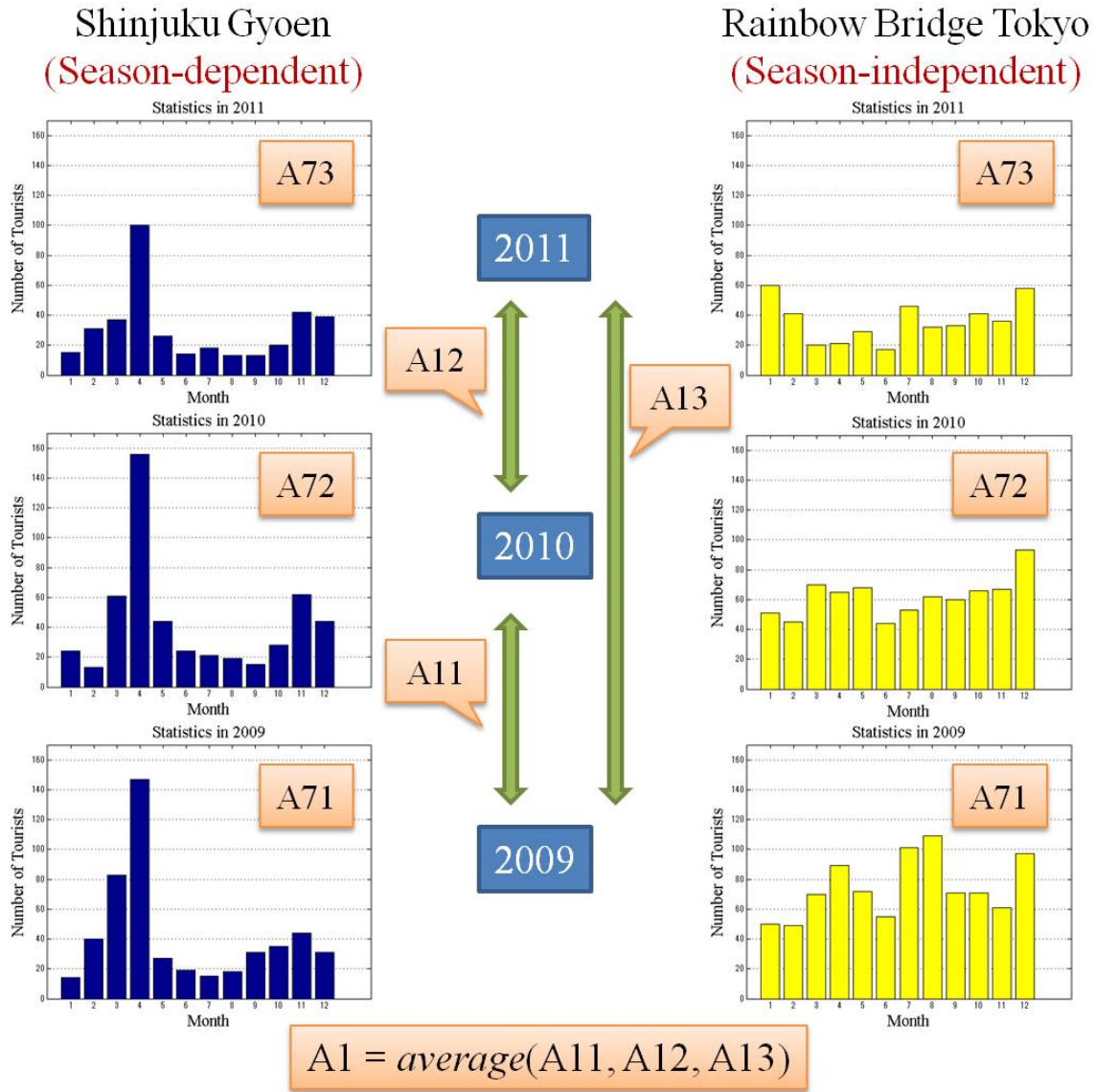


Fig. 5.4 Illustration of attributes 1 and 7 for Shinjuku Gyoen and Rainbow Bridge Tokyo in 2009, 2010 and 2011.

Fig. illustrates the attribute 1 and 7 calculated from the statistics about tourists of Shinjuku Gyoen and Rainbow Bridge Tokyo in 2009, 2010, and 2011. The attribute 2-6 calculated for Shinjuku Gyoen during 2009 to 2011 is shown in Fig. 5.5. In order to calculate attributes A1, A2, A5, and A8 based on the distribution of each month, the number of tourists in each month is normalized by dividing the value by total number of tourists in that year. These attributes represent monthly and yearly variation in the number of tourists. It is expected that tourists want to visit season-dependent spots in specific months such as peak month, whereas season-independent spots would not have such a specific month.

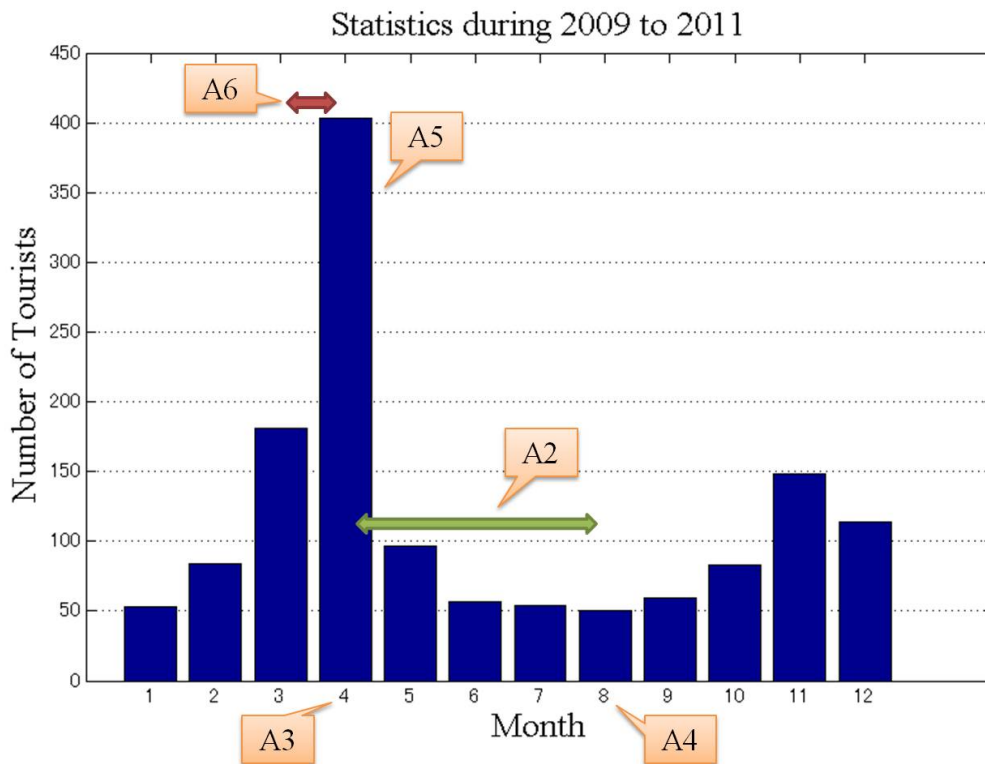


Fig. 5.5 Illustration of attributes 2-6 for Shinjuku Gyoen during 2009 to 2011.

Attributes A3, A4, and A6 relate with the months having many / less number of tourists. These are based on the assumption that different spots in Japan will have a peak of the number of tourists at similar months if such a peak relates with season. For example, if a sightseeing spot is famous for cherry blossoms, it will attract many tourists around April.

Preliminary analysis of the collected data has revealed that the total number of tourists in season-dependent spots tends to be smaller than average of all spots. Considering this observation, attributes A7 are employed as features. It is noted that all of these attributes can be calculated only from metadata of a sightseeing spots, i.e. without downloading actual images. Flickr provides The App Garden<sup>7</sup> that includes numerous API methods available for non-commercial use by developers. The API kits can support various program languages such as Java and PHP. The metadata of images such as shooting timestamps,

<sup>7</sup> <http://www.flickr.com/services/api/>

geotage, and the number of user can be retrieved with XML or JSON format by using these API methods. It is noted that many photos could be uploaded by the same user. Therefore, multiple photos uploaded by the same user will be considered as one user for calculating the number of tourists within the same month.

A classifier is learned based on these 10 attributes. This chapter compares several common machine learning algorithms, of which results are shown in Sec. 5.5.



### 5.3. Image Processing in the Second Phase

The season-dependent scenery attracts tourists owing to the changing color of trees or flowers, but season-independent one doesn't have such season-dependent variation as mentioned above. Therefore, the color feature of images between different months is useful for season classification. By applying content-based classification processing after the first stage in Sec. 5.2, precision is expected to be improved.

In order to extract color feature from specific region, i.e. the dilated edge region which can obtain the color of trees or flowers to reflect the season, image segmentation is employed. Furthermore, the HSV color space is used because human vision can distinguish different hues easily [25].

Only the season-dependent sightseeing spots which were classified in the machine learning phase (Sec. 5.2) are considered in this phase. In order to reduce the computational complexity, we focus on season months, which are defined as every 3 months in a year including a peak month. For example, January, April, July, and October are season months if April is the peak month. Photos uploaded during such season months are downloaded from Flickr. The following six steps are applied for further classification. The photos of single candidate spot are used as input images for these procedures. Fig. 5.6 shows each step of the processing flow and example images of Shinjuku Gyoen.

Step (1) The hue component of HSV is extracted from edge regions, which are segmented by Canny edge detection [18] and then dilated with  $5 \times 5$  kernel by using a morphology operation. The extracted hue component is represented as histogram, of which the number of bins is set to 32.

Step (2) For each of season months, the L1 distance is calculated between each pair of images. The images are considered as neighbor each other if their distance is smaller than average distance of all pairs in that month. The image that has the most neighbor images is selected together with its neighbors.

Step (3) Only the selected images in step 2 are considered in this step. The L1 distance is calculated for all image pairs, and the images of which distance is smaller than average distance of all images are considered neighbors each other. Only the images having fewer neighbors than average are selected in this step.

Step (4) Using the images which are selected in step 3, four centroids are calculated for each of season months.

Step (5) Average distance between a centroid and images in each of season months ( $x$ ) is calculated, which is called intra-cluster distance ( $D_{intra}(x)$ ). As a result, 4 intra-cluster distances are obtained from season months.

Step (6) Distance between centroids of different months is calculated, which is called inter-cluster distance. Let  $D_{inter}(x, y)$  represents distance between centroids of month  $x$  and  $y$ . If at least one pair ( $x, y$ ) of season months satisfies the condition that  $D_{inter}(x, y)$  is greater than  $\max(D_{intra}(x), D_{intra}(y))$ , this spot is classified as season-dependent. Otherwise, it is season-independent spot.

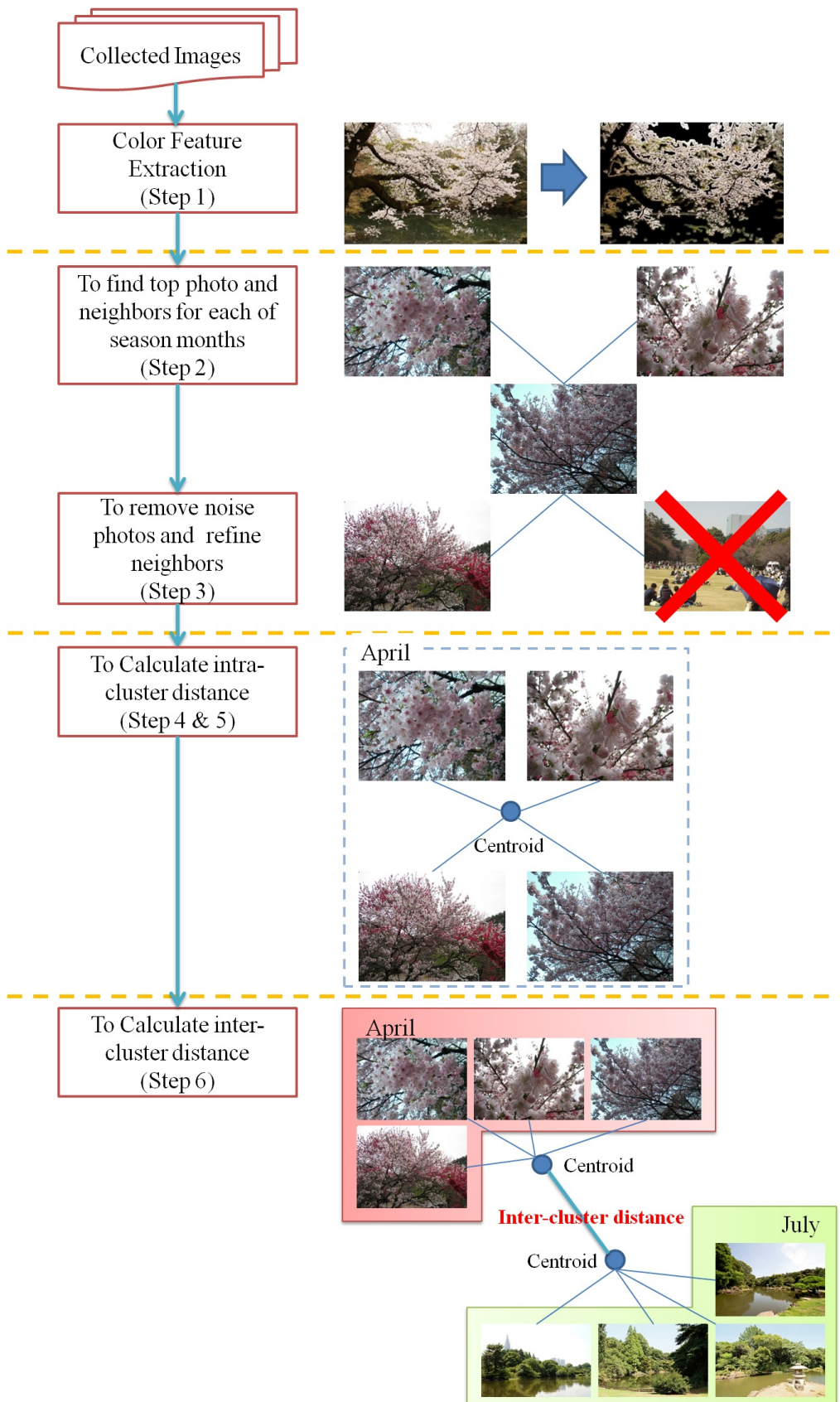


Fig. 5.6 Image processing flow in the second phase.

## 5.4. Experiments

Experiments are conducted in order to evaluate the performance of proposed method. Sightseeing spots to be tested are selected from the sightseeing spot list available on the website of Japan National Tourism Organization (JNTO)<sup>8</sup>. There are total 1,576 spots listed on the website, among which 80 spots, for each of which more than 3,000 images exist on Flickr, were selected manually. Some spots found on the JNTO are famous artificial events such as Fuji Rock Festival, and Jidai Matsuri. As such spots are not suitable for our research purpose, those were manually removed.

After selection, the candidate spots should be defined as season dependent or independent for evaluation. The following procedures are applied when judging season-dependent / independent.

- Use exact text to search on Flickr. This means the spot name with double quotation marks such as “Ueno Park” is used for searching.
- Examine first 100 images sorted by interesting option on Flickr.
- The spot that contains more than 10 images corresponding to one season is considered as season dependent.

In order to explain the third procedure, Fig. 5.7 and Fig. 5.8 show the first page of search result for Ueno Park and Roppongi Hills on Flickr respectively. The searching results are sorted with using interesting option. In Fig. 5.7, 14 photos, which are marked as red rectangles, contain cherry blossoms which are typical objects in spring. On the other hand, there is no photo containing such typical objects corresponding to a certain season in Fig. 5.8. Therefore, Ueno Park and Roppongi Hills are labeled as season-dependent and season-independent respectively.

Table 5.1 shows the summary of test dataset. Among 80 spots, 30 spots are labeled as season-dependent and 50 spots are labeled as season-independent.

---

<sup>8</sup> <http://www.jnto.go.jp/eng/location/maps/>

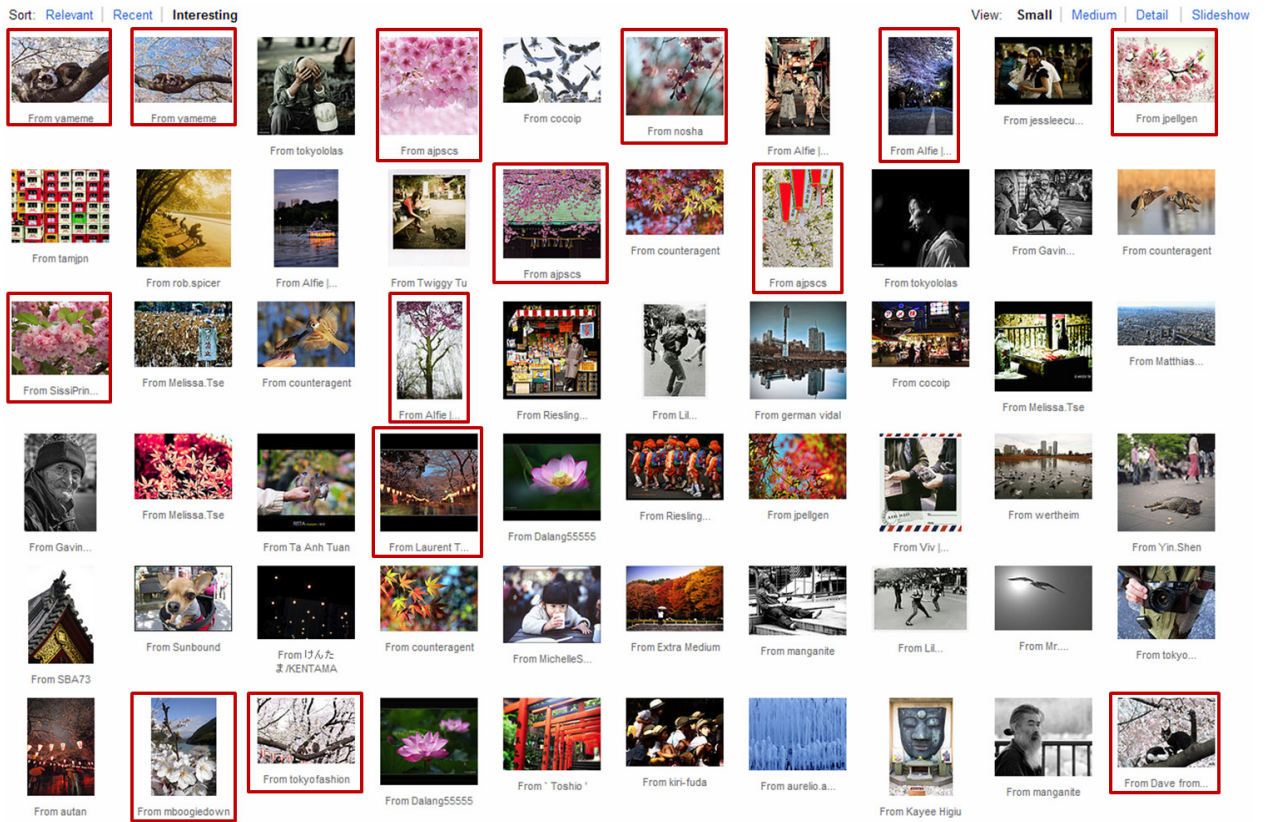


Fig. 5.7 Photos of Ueno Park collected from Flickr.



Table 5.1 Summary of test dataset.

Label	Name of Sightseeing Spots				
Season-dependent	Arashiyama Kyoto	Byodo-in	Chion-in	Gosho Kyoto	Hakuba Japan
	Heian Shrine	Himeji Castle	Inokashira Park	Karuizawa	Kenrokuen
	Kinkakuji	Kiyomizudera	Koishikawa-Korakuen	Korakuen	Lake Kawaguchi
	Matsumoto Castle	Nagoya Castle	Nanzenji	Nara Park	Niseko
	Odawara Castle	Osaka Castle	Rikugien	Sankeien	Shinjuku Gyoen
	Shirakawago	Takaosan	Tenryuji	Tokyo University	Ueno park
Season-independent	21st Century Museum of Contemporary Art Kanazawa	Akihabara	Aquarium Kaiyukan Osaka	Atomic Bomb Dome Hiroshima	Churaumi Aquarium Okinawa
	Daibutsu Kamakura	Dotonbori	Fushimi Inari taisha	Ghibli Museum	Haneda Airport
	Hiroshima Castle	Kabukicho	Karatsu	Kasuga taisha	Kumamoto Castle
	Kurashiki	Kyoto Station	Landmark Tower Yokohama	Makuhari Messe	Meiji Shrine
	Miyajima	Mt. Fuji	Narita Airport	National Art Center Tokyo	Nijo Castle
	Onomichi	Open Air Museum Hakone	Peace Memorial Park Hiroshima	Rainbow Bridge Tokyo	Roppongi Hills
	Sakuragicho	Sanrio Puroland	Sensoji	Shimabara	Shinsekai
	Shuri Castle	Todaiji	Tokyo Disneyland	Tokyo Dome	Tokyo International Forum
	Tokyo Midtown	Tokyo Sky Tree	Tokyo Tower	Tsukiji Market	Ueno Zoo
	Universal Studios Japan	Unzen Japan	Yamashita Park	Yanagawa	Yasaka Shrine



### 5.4.1. Experimental Results of the First Phase

The test datasets are processed with a data mining software WEKA [26]. In order to evaluate the performance of the proposed method, combination of 3 sets of attributes and 4 machine learning algorithms are examined in the first phase. The 10-fold cross-validation is used to evaluate the performance of classifiers. The sets of attributes used for the experiment are as follows:

- Tourists: An attribute set consisting of 10 attributes calculated from the number of tourists within 2009 to 2011. This set is the same as proposed in Sec. 5.1.
- Photos: An attribute set consisting of 10 attributes calculated from the number of photos within 2009 to 2011, by replacing the number of tourists with that of photos in the definitions in Sec. 5.1.
- Mix: A combination set of tourists and photos. This set contains 20 attributes.

The following 4 classifiers are employed for comparing the results of classification.

- Naive Bayes: A Naive Bayes classifier using estimator classes [43].
- LibSVM: A library for Support Vector Machines [27, 28].
- JRip: A propositional rule learner, Repeated Incremental Pruning to Produce Error Reduction (RIPPER) [44].
- J48: A clone of the C4.5 decision tree learner [45].

Table 5.2 Comparison of different attribute sets (without discretization) applied on 4 classifiers.

Attribute Set	Classifier	Precision (%)	Recall (%)	F-measure (%)
Tourists	Naive Bayes	71.4	83.3 *	76.9 *
	LibSVM	0	0	0
	JRip	73.1 *	63.3	67.9
	J48	71.9	76.7	74.2
Photos	Naive Bayes	38.7	80.0	52.2
	LibSVM	0	0	0
	JRip	63.6	46.7	53.8
	J48	40.0	6.70	11.4
Mix	Naive Bayes	51.0	83.3 *	63.3
	LibSVM	0	0	0
	JRip	71.4	66.7	69.0
	J48	71.9	76.7	74.2



Table 5.2 shows the experimental result in the first phase. It is seen that Naive Bayes with attribute set of Tourists can get the best result in f-measure and recall. The highest score is marked with an asterisk (\*). Although precision of Naive Bayes is lower than JRip and J48, recall is higher. As irrelevant spots can be discriminated in the second phase, the first phase can assign higher priority to recall than precision. Therefore, attribute set of Tourists with Naive Bayes is the best combination in the first phase.

As discretization is often used as preprocessing for data mining, another experiment is conducted, in which all attributes except A3 and A4 are discretized. The result is shown in Table 5.3.

Table 5.3 shows the improvement of performance by libSVM compared with the result shown in Table 5.2, in which both of precision and recall are 0. However, its recall and F-measure are much worse than Naive Bayes' ones. The experiment of applying Naive Bayes with attribute set of tourists can still obtain the best result in recall and F-measure.

Table 5.3 Comparison of different attribute sets (with discretization) applied on 4 classifiers.

Attribute Set	Classifier	Precision (%)	Recall (%)	F-measure (%)
Tourists	Naive Bayes	75.0	70.0 *	72.4 *
	LibSVM	81.8 *	30.0	43.9
	JRip	50.0	60.0	54.5
	J48	57.6	63.3	60.3
Photos	Naive Bayes	50.0	43.3	46.4
	LibSVM	55.6	16.7	25.6
	JRip	57.9	36.7	44.9
	J48	65.0	43.3	53.0
Mix	Naive Bayes	72.4	70.0 *	71.2
	LibSVM	10.0	6.7	12.5
	JRip	54.5	40.0	46.2
	J48	59.4	63.3	61.3

#### 5.4.2. Experimental Results of the Second Phase

In the second phase, the test dataset contains 35 sightseeing spots which are classified as season-dependent spot in the first phase. In order to evaluate the proposed method in the second phase, color features extracted from different regions as well as different calculation of  $D_{intra}(x)$  in step 5 of Sec. 5.3 are calculated for comparison.

- Hue & Avg. Distance of All: Hue component is extracted from whole image. Average

distance of all image pairs for each season month is used as  $D_{intra}(x)$ .

- Hue & Avg. Distance of Centroid: Hue component is extracted from whole image. The  $D_{intra}(x)$  is calculated as proposed in Sec. 5.3.
- Hue on Edge & Avg. Distance of All: Hue component is extracted from edge region. Average distance of all image pairs for each season month is used as  $D_{intra}(x)$ .
- Hue on Edge & Avg. Distance of Centroid (Proposed method): Hue component is extracted from edge region. The  $D_{intra}(x)$  is calculated as proposed in Sec. 5.3.

Table 5.4 Comparison of different image processing method.

Method	Precision (%)	Recall (%)	F-measure (%)
Hue & Avg. Distance of All	66.7	32.0	43.3
Hue & Avg. Distance of Centroid	74.1	80.0	76.9
Hue on Edge & Avg. Distance of All	76.9	40.0	52.6
Hue on Edge & Avg. Distance of Centroid	75.9	88.0	81.5 *

Table 5.4 shows the experimental result in the second phase. The comparison of color features extracted from different region shows that hue component extracted from edge region performs better because the whole image contains many noises. On the other hand, typically seasonal objects such as flowers or leaves can be segmented by edge detection. It is also observed that using the proposed definition of  $D_{intra}(x)$  can get better result. The Avg. Distance of All is easily influenced by extreme image that is not similar to others in the same season month. Therefore, the proposed method, i.e. Hue on Edge & Avg. Distance of Centroid can get the best result in this phase.

Table 5.5 Performace of proposed method including 1st and 2nd phase.

Method	Precision (%)	Recall (%)	F-measure (%)	Correctly Classified Rate (%)
1st Phase	71.4	83.3	76.9	81.3
1st & 2nd Phase	75.9	73.3	74.6	81.3

Table 5.5 compares the best result in the first phase and total result including the first

and second phases. It is seen that by applying the second phase, precision increases but recall decreases while keeping correctly-classified rate. That is, the number of false-negative increases but that of false-positive decreases. This result indicates the second phase contributes to eliminate season-dependent spots misclassified at the first phase.

## 6. Conclusions and Future Research Directions

This thesis proposes image classification methods targeting sightseeing spots of various situations. Situation-oriented grouping of sightseeing images is useful for tourists planning when to visit which sightseeing spots as noted in Chapter 1. Although image classification / annotation methods for general-purpose have recently been studied by many researchers, the contribution of the thesis is that efficient image classification method is established based on organization of various kinds of situations and consideration on various features and those integration in terms of characteristics of each situation.

The situations handled in this thesis are classified into weather-related, time-related, and season-related types based on the characteristics of target situations. For weather-related situations such as sunny and cloudy, color features of sky regions are utilized for classification. Time-related situations include night-time, daytime, and sunrise/sunset, each of which corresponds to different times of the day. On the other hand, in the case of season-related situations, whether or not scenery will change according to the situation depends on the characteristics of a sightseeing spot. Therefore, a preprocessing for classifying various sightseeing spots into season-dependent and season-independent spots is proposed.

A content-based image classification method is proposed for weather-related and time-related situations. The images are classified in a hierarchical manner. In each stage, the extraction of local color features is performed based on the composition of an image and typical colors in target situations. The proposed method can obtain high precision and recall for classifying images into target situations as shown in experimental results.

The metadata information attached to images such as timestamps and geotag are employed for time-related situations to complement the content-based image classification. In order to increase the accuracy of the classification, the time windows, which can be adjusted according to the geolocation of sightseeing spots, are proposed as filters to verify the clusters obtained by content-based approach. Experimental results show that this hybrid approach can improve precision while maintaining recall in most cases.

For season-related situations, a two-stage classification method is proposed for classifying sightseeing spots into season-dependent and season-independent ones as preprocessing of image classification. In order to reduce the cost of image processing and network load for downloading photos of target sightseeing spots, the statistical features of

sightseeing spots are calculated using metadata only, based on which classifier is trained in the first stage. In the second stage, image processing is only applied to the spots classified as season-dependent in the first stage for improving the classification accuracy. Experimental results show that the proposed method can classify actual sightseeing spots with high precision and recall.

Regarding the contribution of this thesis, the proposed image classification methods can realize advanced web-based tourist services as noted in Chapter 1. Tourism is one of important and promising industries for many countries, because tens of thousands of tourists can bring considerable income to the countries. With the development of tourist services, people tend to utilize these services when making plan, searching information or route planning on map, and sharing experiences and photos with friends. Therefore, the results of this thesis are meant to contribute to tourism and related applications such as recommendation services or auto-annotation on web albums. Furthermore, as the volume of images and metadata available on the Web is still increasing at a rapid rate, the contribution of the thesis may have numerous other applications.

## References

- [1] 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, No. 1-3, pp. 193-233, 2004.
- [2] A. K. Jain and A. Vailaya, "Image Retrieval using Color and Shape," *Pattern Recognition*, Vol.29, No. 8, pp. 1233-1244, 1996.
- [3] C. Carson, S. Belongie, H. Greenspan, and J. Malik, "Blobworld: Image Segmentation using Expectation-Maximization and Its Application to Image Querying," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, Vol.24, No. 8, pp. 1026-1038, 2002.
- [4] A. B. Dahl and H. Aanaes, "Effective Image Database Search via Dimensionality Reduction," *IEEE Computer Conf. on Computer Vision and Pattern Recognition*, pp. 1-6, 2008.
- [5] R. C. Veltkamp and M. Tanase, "A Survey of Content-Based Image Retrieval Systems," *Multimedia Systems and Applications Series*, Vol.21, pp. 47-101, 2002.
- [6] A. Vailaya, M. Figueiredo, A. Jain, and H. J. Zhang, "Image Classification for Content-Based Indexing," *IEEE Trans. on Image Processing*, Vol.10, No. 1, pp. 117-130, 2001.
- [7] A. Vailaya, M. Figueiredo, A. Jain, and H. J. Zhang, "Content-based hierarchical classification of vacation images," *IEEE Int. Conf. on Multimedia Computing and Systems*, Vol.1, pp. 518-523, 1999.
- [8] A. Vailaya, M. Figueiredo, A. Jain, and H. J. Zhang, "A Bayesian Framework for Semantic Classification of Outdoor Vacation Images," in *Proc. SPIE Storage Retrieval Image Video Databases VII*, Vol. 3656, pp. 415-426, 1999.
- [9] 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.
- [10] S. Silakari, M. Motwani, and M. Maheshwari, "Color Image Clustering using Block Truncation Algorithm," *Int. J. of Computer Science Issues (IJCSI)*, Vol.4, No. 2, pp. 31-35, 2009.
- [11] A. Sleit, A. L. A. Dalhoum, M. Qataweh, M. Al-Sharief, R. Al-Jabaly, and O.

- Karajeh, "Image Clustering using Color, Texture and Shape Features," *KSII Trans. on Internet and Information Systems*, Vol.5, No. 1, pp. 211-227, 2011.
- [12] 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.
- [13] J. Hartigan and M. Wong, "Algorithm as 136: A K-means Clustering Algorithm," *J. of the Royal Statistical Society, Series C (Applied Statistics)*, Vol.28, No. 1, pp. 100-108, 1979.
- [14] J. Hartigan, *Clustering Algorithms*, JohnWiley & Sons, Inc., New York, 1975.
- [15] N. Zhou, W. M. Dong, J. X. Wang, and P. Jean-Claude, "Simulating Human Visual Perception in Nighttime Illumination," *Tsinghua Science & Technology*, Vol.14, No. 1, pp. 133-138, 2009.
- [16] L. G. Liu, R. J. Chen, L. Wolf, and D. Cohen-Or, "Optimizing Photo Composition," *Computer Graphics Forum*, Vol.29, pp. 469- 478, 2010.
- [17] Rule of thirds (Wikipedia), [http://en.wikipedia.org/wiki/Rule of thirds](http://en.wikipedia.org/wiki/Rule_of_thirds), accessed on September 26, 2011.
- [18] J. F. Canny, "A Computational Approach to Edge Detection," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, Vol.8, No. 6, pp. 679-698, 1986.
- [19] N. Otsu, "A Threshold Selection Method from Gray-Level Histograms," *IEEE Trans. on Systems, Man, and Cybernetics*, Vol.9, No. 1, pp. 62-66, 1979.
- [20] S. Papadopoulos, C. Zigkolis, G. Toliass, Y. Kalantidis, P. Mylonas, Y. Kompatsiaris, and A. Vakali, "Image Clustering through Community Detection on Hybrid Image Similarity Graphs," *IEEE International Conference on Image Processing (ICIP)*, pp. 2353-2356, 2010.
- [21] P. A. Moellic, J. E. Haugeard, and G. Pittel, "Image clustering based on a shared nearest neighbors approach for tagged collections," *Proceedings of the 2008 international conference on Content-based image and video retrieval*, pp. 269-278, 2008.
- [22] D. Buhalis and R. Law, "Progress in information technology and tourism management: 20 years on and 10 years after the Internet—The state of eTourism research," *Tourism Management* Volume 29, Issue 4, pp. 609-623, 2008.
- [23] R. Law, S. Qi, and D. Buhalis, "Progress in tourism management: a review of website evaluation in tourism research," *Tourism Management*, Volume 31, Issue 3, pp. 297-313, 2010.

- [24] J. P. Lucas, N. Luz, M. N. Moreno, R. Anacleto, A. A. Figueiredo, and C. Martins, "A hybrid recommendation approach for a tourism system, Expert Systems with Applications," Vol. 40, Issue 9, pp. 3532-3550, 2013.
- [25] H. D. Cheng, X. H. Jiang, Y. Sun, and J. Wang, "Color image segmentation: advances and prospects," Pattern Recognition, Vol. 34, Issue 12, pp. 2259-2281, 2001.
- [26] M. Hall, E. Frank, G. Holmes, B. Pfahringer, P. Reutemann, I. H. Witten, "The WEKA Data Mining Software: An Update, SIGKDD Explorations," Vol. 11, Issue 1, pp. 10-18, 2009.
- [27] Y. EL-Manzalawy and V. Honavar, "WLSVM : Integrating LibSVM into Weka Environment," Software available at <http://www.cs.iastate.edu/~yasser/wlsvm>, 2005.
- [28] C. C. Chang and C. J. Lin, "LIBSVM : a library for support vector machines," ACM Transactions on Intelligent Systems and Technology, Vol. 2, No.3, 2011.
- [29] Y. Li, D. J. Crandall, and D. P. Huttenlocher, "Landmark Classification in Large-scale Image Collections," IEEE 12th International Conference on Computer Vision, pp. 1957-1964, 2009.
- [30] T. Quack, B. Leibe, and L. V. Gool, "World-scale mining of objects and events from community photo collections," Proceedings of the 2008 international conference on Content-based image and video retrieval, pp. 47-56, 2008.
- [31] W. Zhou, Y. Lu, H. Li, and Q. Tian, "Canonical Image Selection by Visual Context Learning," 20th International Conference on Pattern Recognition, pp. 834-837, 2010.
- [32] L. Yang, J. Johnstone, and C. Zhang, "Ranking canonical views for tourist attractions," Multimedia Tools and Applications, Vol. 46, Issue 2-3, pp. 573-589, 2010.
- [33] L. Kennedy and M. Naaman, "Generating Diverse and Representative Image Search Results for Landmarks," Proceedings of the 17th international conference on World Wide Web, pp. 297-306, 2008.
- [34] Y. Jing, S. Baluja, and H. Rowley, "Canonical Image Selection from the Web," Proceedings of the 6th ACM international conference on Image and video retrieval, pp. 280-287, 2007.
- [35] D. G. Lowe, "Object recognition from local scale-invariant features," International Conference on Computer Vision, pp. 1150-1157, 1999.
- [36] N. Sharda, Tourism Informatics. Visual Travel Recommender Systems, Social Communities and User Interface Design, Information Science Reference, 2009.



- [37] L. White, "Facebook, friends and photos: A snapshot into social networking for generating travel ideas," In: Sharda, N. (ed.) *Tourism Informatics: Visual Travel Recommender Systems, Social Communities, and User Interface Design*, Information Science Reference, pp. 115–129, 2010.
- [38] H. Bay, A. Ess, T. Tuytelaars, and L. V. Gool, "SURF: Speeded Up Robust Features," *Computer Vision and Image Understanding (CVIU)*, Vol. 110, No. 3, pp. 346-359, 2008.
- [39] Image Retrieval in CLEF (ImageCLEF), Available: <http://www.imageclef.org/>.
- [40] B. Thomee, A. Popescu, "Overview of the ImageCLEF 2012 Flickr Photo Annotation and Retrieval Task," In: *CLEF 2012 working notes*, 2012.
- [41] S. Nowak, K. Nagel and J. Liebetrau, "The CLEF 2011 Photo Annotation and Concept-based Retrieval Tasks," In: *CLEF 2011 working notes*, 2011.
- [42] S. Nowak and M. Huiskes, "New Strategies for Image Annotation: Overview of the Photo Annotation Task at ImageCLEF 2010," In: *CLEF 2010 working notes*, 2010.
- [43] G. H. John and P. Langley, "Estimating Continuous Distributions in Bayesian Classifiers," *Proceedings of the Eleventh Conference on Uncertainty in Artificial Intelligence*, pp. 338-345, 1995.
- [44] W. W. Cohen, "Fast Effective Rule Induction," In: *Twelfth International Conference on Machine Learning*, pp. 115-123, 1995.
- [45] J. R. Quinlan, *C4.5: Programs for Machine Learning*, Morgan Kaufmann Publishers, Inc., San Mateo, California, 1993.



## **Related Publications**

### **Journal papers**

1. Chia-Huang Chen, Yasufumi Takama, "Situation-Oriented Hierarchical Classification for Sightseeing Images Based on Local Color Feature," *Journal of Advanced Computational Intelligence and Intelligent Informatics*, Vol. 17, No. 3, pp. 459-468, 2013.
2. Chia-Huang Chen, Yasufumi Takama, "Hybrid Approach of Situation-Oriented Classification of Sightseeing Spot Images Based on Visual and Tag Information," *Journal of Information Science and Engineering*, Accepted, 2013.

### **International Conference papers**

1. Chia-Huang Chen, Yasufumi Takama, "Situation-Oriented Clustering of Sightseeing Spot Images Using Visual and Tag Information," *Joint 6th International Conference on Soft Computing and Intelligent Systems (SCIS) and 13th International Symposium on Advanced Intelligent Systems (ISIS)*, pp. 416-421, 2012.
2. Chia-Huang Chen, Yasufumi Takama, "Hybrid Approach of Using Visual and Tag Information for Situation-Oriented Clustering of Sightseeing Spot Images," *2012 Conference on Technologies and Applications of Artificial Intelligence (TAAI 2012)*, pp. 256-261, 2012.
3. Chia-Huang Chen, Yasufumi Takama, "Classification of Season-Dependent Sightseeing Spots using Statistical Features Obtained from Metadata," *The Third International Workshop on Advanced Computational Intelligence and Intelligent Informatics (IWACIII2013)*, Accepted, 2013.

### **Other papers**

1. Chia-Huang Chen, Lieu-Hen Chen, Yasufumi Takama, "Proposal of Situation-based Clustering of Sightseeing Spot Images based on ROI-based Color Feature Extraction," *The 26th Annual Conference of the Japanese Society for Artificial Intelligence (JSAI2012)*, No. 4M1-IOS-3c-3, 2012.

## **Acknowledgments**

First of all, I am sincerely grateful to my supervisor, Prof. Yasufumi Takama, for teaching and training me to become a researcher and finish this thesis. He supported me not only about research but also encourage me when I feel stressful and help me a lot for my living in Japan in these three years. I have learned the spirit of good working attitude from him.

Next, I am very grateful to all the members of my doctoral thesis committee, Prof. Toru Yamaguchi, Prof. Hiroshi Ishikawa, and Prof. Kaoru Hirota, for their useful comments and helpful suggestions that have significantly improved the quality of my thesis.

I also would like to thank to the members of Takama Laboratory and Yamaguchi Laboratory, especially Prof. Yi-Hsin Ho, Mr. Shunichi Hattori and Mr. Zhongjie Mao, for their helping to support my living and solve many problems about Japanese language.

Finally, I would like to give my greatest thanks to my parents, my grandmother, my elder brother, my family, and Miss Wen-Ling Wang for their care and always talk with me, let me feel I am not along in a foreign country and give me more energy to face and solve any problems.

My doctorate research was fully financially supported by the Asian Human Resources Fund Scholarship, generously provided by Tokyo Metropolitan Government.