

# Image Retrieval based on Object Extraction and *Kansei* Estimation

Nanik Suciati <sup>† ‡</sup>, Darlis Herumurti <sup>†</sup>, Joko Lianto Buliali <sup>†</sup>,  
Dyah Wardhani Kusuma <sup>†</sup>, Ahmad Saikhu <sup>†</sup>, Chie Muraki Asano <sup>††</sup>, and Akira Asano <sup>‡</sup>

<sup>†</sup> Department of Informatics, Faculty of Information Technology,  
Sepuluh Nopember Institute of Technology (ITS), Sukolilo, Surabaya 60111, Indonesia

<sup>††</sup> International Student Center, Hiroshima University,  
Higashi-Hiroshima, Hiroshima 739-8524, Japan

<sup>‡</sup> Department of Information Engineering, Graduate School of Engineering,  
Hiroshima University, Higashi-Hiroshima, Hiroshima 739-8521, Japan

E-mail: <sup>†</sup> {nanik, darlis, joko, dyah, saikhu}@its-sby.edu, <sup>††</sup> chiem@mikeneko.jp,  
<sup>‡</sup> asano@mis.hiroshima-u.ac.jp, nanik@hiroshima-u.ac.jp

**Abstract** This research proposes an image retrieval system based on object extraction and *Kansei* (that is, a high-order function of the brain, including inspiration, intuition, pleasure and pain, taste, curiosity, aesthetics, emotion, sensitivity, attachment and creativity) estimation, which enables retrieving images by using object names and/or human *Kansei* as retrieval keywords. The proposed system focuses on the extracting objects from image and the relationships between objects and human *Kansei*. It depicts more clearly how human impression causes from characteristics of the extracted objects in images, whereas other methods evaluate impressions from characteristics of the whole or parts of images. Some experimental results in object recognition, human *Kansei* estimation and image retrieval are presented.

**Keywords** Image retrieval system, object recognition, human *Kansei* estimation

## 1. Introduction

In recent years, content-based image retrieval (CBIR) has become increasingly popular with the rapid development of storage devices and digital photographs. CBIR systems that can automatically organize image database have been developed both in research laboratories and for commercial concerns. Most of the early researches in CBIR systems perform image retrieval based primarily on low-level image features, such as color and texture. The experiments with these systems have shown that two images which have similar low level features does not ensure that they will also have similar high-level semantic content. In order to provide adequate image retrieval, it is necessary to organize the image database by using high-level content, such as, object content.

Researches of object recognition for automated indexing of large image databases have been reported. One approach is to segment images into regions whose features are typical of certain well-known objects, such as tigers and zebras, which have characteristic colors and texture [1]. Local structural features called *consistent line clusters* are used to recognize man-made objects in images [2]. Another approach motivated by the recent availability of large annotated image, which learns correlations between visual elements in image and words, is used to solve the object recognition problem [3].

Although the object based image retrieval is promising, it provides only a partial solution to the retrieval problem, because users might want to retrieve an image not based on the objects in the image, but based on their particular “feeling” to the image. It evokes much attention in developing

*Kansei* image retrieval. A new approach applied with some methods that introduce *Kansei* to multimedia is studied by discovering the most attractive regions and features of images [4]. Other approach proposes *Kansei* method as the means to join psychology, which deals with concepts, and computer science, which deals with physical measurable phenomena such as color, direction, movement, or position [5]. In addition, a different approach recognizes sky/earth/water regions from scenery image, and then describes more detail objects and the impression words from recognized regions by using neural network [6].

To provide an alternative to the needless solution of the retrieval problem, the goal of this research is to develop an image retrieval system, which can organize the image database based on objects and human *Kansei* (here, it means impression words) automatically. Object recognition is implemented by computing Euclidian distance between features vector of each region of test image resulted from the *JSEG* segmentation method [7] and the centroids of features vectors clusters resulted from training stage, in order to find the best suitable cluster representing object name. The recognized objects are furthermore used to estimate the impression words. Whereas other methods evaluate impression words by characteristics of the whole or parts of images, this research focuses on describing the relationships between characteristics of objects in image and human *Kansei*. It depicts more clearly how human impression is caused from characteristics of the extracted objects in images. This system can be extended to recognize more objects, and estimate more impression words.

## 2. The developed system

The developed system is an image retrieval system, which automatically organizes images based on objects and impression words content. Fig. 1 shows an overview of the developed system. Two important processes in storing image task are object recognition and impression words estimation. First, objects in the image are recognized, and then the characteristics of the recognized objects are used to

estimate the impression words. Finally, the recognized objects and the estimated impression words are used to index the image into image database. In retrieving images task, the user can retrieve images by using object and/or impression words as keywords.

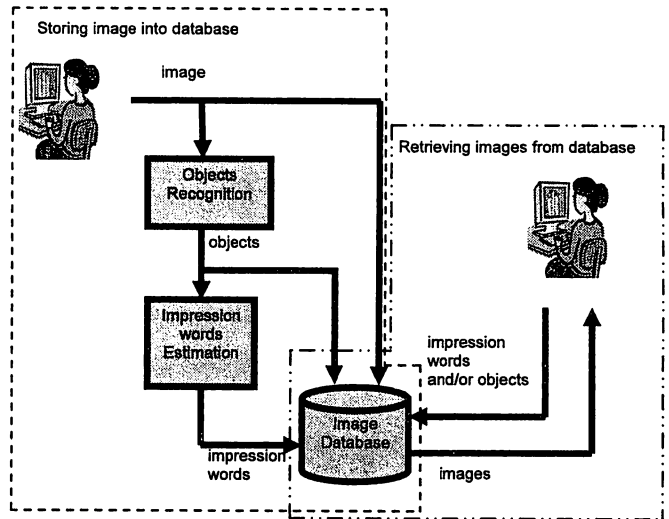


Fig.1. Overview of the developed system.

### 2.1. Object recognition process

Fig. 2 shows an overview of object recognition process, which consist of training and recognition stage. The objective of training stage is to determine  $k$ -clusters of feature vectors, which furthermore are used in recognition stage. Each cluster represents one object name. In training stage, training images are segmented into some regions using *JSEG* segmentation method [7]. Here, extracting 9 features and describing object name for each region are carried out in order to get feature vectors. The 9 features extracted from each region are size, mean of each color component in *RGB* space, standard deviation of each component in *RGB* space, center of gravity in  $x$  and  $y$  component. The 12 different objects described manually for each region are bright sky, dark sky, green grass, dead grass, colorful grass, land, asphalt road, snow, water, stone, mountain, and building. Clustering feature vectors using the *k-means clustering* algorithm [8] is applied to group similar feature vectors. Feature vectors, belonging to the same cluster, represent a region having similar characteristics, thus, identifying the same object. By experiment, reducing the value of  $k$  will

also reduce the computation time of object recognition. The value of  $k$  is chosen with the consideration of tradeoff between computation time and performance of recognition. In recognition stage, the nearest cluster for each feature vector of new image is determined by computing *Euclidian distance* between feature vector and the centroid of each cluster. Once determined, the object representing each feature vector of new image can be estimated.

Image segmentation is an important step, because the result of segmentation will influence performance of the object recognition process. The segmentation method should be robust and result object-like regions. For example, human face, zebra or tiger, should be segmented into one region. Those are reasons why *JSEG* [7], a robust, unsupervised segmentation method based on color and texture features is chosen. *JSEG* algorithm consists of two stages, that is color quantization and spatial segmentation. At color quantization stage, colors in the image are quantized to several representing classes that can be used to differentiate regions in the image. Then, image pixel colors are replaced by their corresponding color class label, thus forming a class-map of the image. At spatial segmentation stage, a

criterion for “good” segmentation is calculated using the formed class-map. Applying the criterion to local windows in the class-map results “J-Image”, in which high and low values correspond to possible region boundaries and region centers, respectively. A region growing method is then used to segment the image based on the multi-scale J-Images.

*K-means clustering* is one of the simplest unsupervised learning algorithms that classify a given data set through a certain number of clusters (assume  $k$  clusters) fixed a priori. The main idea is to define  $k$  centroids, one for each cluster. These centroids should be placed in a cunning way because of different location causes different result. The better choice is to place them far away from each other as possible. The next step is to take each point belonging to a given data set and associate it to the nearest centroid. When no point is pending, the first step is completed and an early grouping is determined. Then, the re-calculation of  $k$  new centroids of clusters resulting from the previous step can be executed. If  $k$  new centroids are different to the previous one, a new grouping should be computed between the same data set points and the nearest new centroid. Looping continues until  $k$  new centroids are same to the previous one.

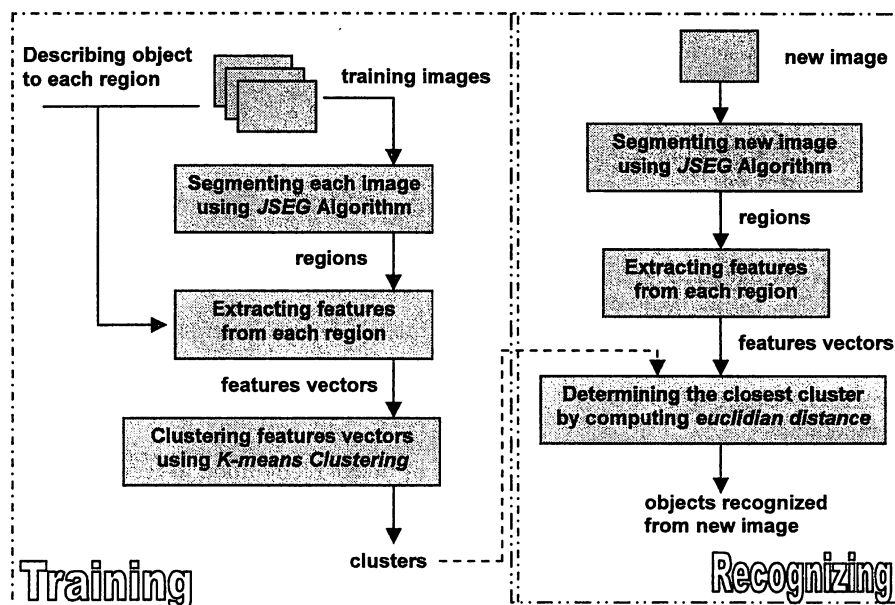


Fig.2. Overview of object recognition process.

The aim of *k-means clustering* is to minimize the objective function

$$J = \sum_{j=1}^k \sum_{i=1}^n \|x_i^{(j)} - c_j\|^2 ; \text{ where } \|x_i^{(j)} - c_j\|^2$$

is a chosen distance between a data point  $x_i^{(j)}$  and the centroids  $c_j$ . The distance is an indicator of the distance of the  $n$  data points from their respective cluster centers.

## 2.2 Impression words estimation process

Whereas other methods evaluate impression words from characteristics of the whole or parts of images, this research focuses on describing the relationships between characteristics of objects in image and human *Kansei*. It depicts more clearly how human impression causes from characteristics of the extracted objects in images. For example, ‘bright’ impression is caused of wide size of bright sky object in image, ‘crowded’ impression is caused of large number of people object in image, and ‘lonely’ impression is caused of a human face that seems to be sad. List of impression words used in this research and the description of sensation or feeling for each word are shown below:

<b>Cold</b>	feeling coldness from something having a very low temperature
<b>Dry</b>	feeling grown by something lacking normal moisture or depleted of water
<b>Fresh</b>	sensation of restored energy grown by purity or pleasant view
<b>Dark</b>	feeling caused of a view having a little light
<b>Bright</b>	feeling caused of a view having lots of light
<b>Warm</b>	sensation of enthusiasm or feeling grown by hotter temperature
<b>Cool</b>	sensation of relax or feeling grown by a refreshingly low temperature
<b>Open</b>	feeling grown by unobstructed view
<b>Crowded</b>	feeling caused of a view contains a large number of things/people positioned or considered together
<b>Pleasant</b>	sensation caused by agreeable/nice stimuli in appearance of human
<b>Lonely</b>	sensation caused by appearance of dejected human from being alone

Observation to the impression words given to some images by several respondents results the relationship be-

tween impression words and characteristics of objects in image. Table 1 shows a list of impression words influenced by characteristics of objects. Once the objects in the image are recognized, estimating the impression words is quite easy. If the characteristics of objects have value more than threshold, then the responding impression word is chosen. Threshold for each impression words is determined based on the response mean of several respondents to the characteristics of objects. For example, threshold for ‘cold’ impression word is determined based on the response mean of several respondents to some images with different size of snow.

Table 1. Relationship between impression words and characteristics of objects.

Impression word	Characteristics of objects
Cold	Size of snow
Dry	Size of dead grass, stone, land, asphalt road
Fresh	Size of water, colorful grass
Dark	Size of dark sky
Bright	Size of bright sky
Warm	Size of colorful grass
Cool	Size of green grass, mountain
Open	Size of bright sky, dark sky
Crowded	Size of human face, car Number of human face, car
Pleasant	Size of human face Expression of human face
Lonely	Size of human face Expression of human face

Performance of impression words estimation process depends ultimately on the performance of object recognition. If object recognition process can recognize what objects in image accurately, then impression words estimation process will give accurate results based on the relationships table. The difficulty in our system is the low performance of object recognition to recognize each human face, due the lack of segmentation method that cannot divide each human face in image as separate region. That is the reason to postpone estimation of impression words like ‘crowded’, ‘pleasant’ and ‘lonely’, which involve characteristics of human face. In addition, recognizing expression of human face automatically is also important to estimate ‘pleasant’ and ‘lonely’ impression words.

### 3. Experimental Results

150 images collected from various sources are used in the experiment, 120 as training images and 30 as test images. All images are stored into database together with objects and impression words contents. All images are 8 bit RGB full-color and have maximum size 200 x 200 pixels. Some experiments are carried out to measure performance of object recognition, impression word estimation and image retrieval.

In object recognition, each region resulted from segmentation process is recognized as one object. All different recognized objects will be objects content of an image. Then, the relationship of object characteristics and impression words are used to estimate impression words. Table 2 shows a quantitative measure of object recognition performance for 30 test images. The average of correct recognition is 81%. Bright sky has the highest percentage, whereas building has the lowest percentage. This is a direct impact of using different number of regions representing each object as data training, where the number of bright sky regions is the largest and the number of building regions is the smallest one. Adding the number of data training can improve performance of object recognition.









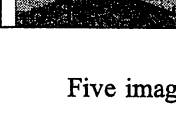
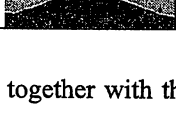
Table 2. Performance of object recognition.

Object Name	Number of regions, which		Percentage of correctness in recognition
	Correctly recognized	Incorrectly recognized	
Bright sky	57	3	95 %
Dark sky	8	2	80 %
Green grass	41	4	91 %
Dead grass	26	4	87 %
Colorful grass	14	6	70 %
Land	16	4	80 %
Asphalt road	13	2	87 %
Snow	8	2	80 %
Water	13	2	87 %
Stone	7	3	70 %
Mountain	12	3	80 %
Building	3	2	60 %
Average			81 %

Table 3 shows the results of object recognition and impression words estimation. The segmentation result shows that the segmentation method not always segments

one object in the image into one region. This characteristic make it difficult to enumerate accurately the number of entity objects like car, human face, building, etc, whereas it is easy to enumerate the total size of objects in image.







Table 3. The results of object recognition and impression words estimation.

Color Image	Segmentation	Object recognition	Impression words estimation
		Bright sky Green grass	Cool Open
		Bright sky Dead grass Green grass	Dry Bright Open
		Snow Building Bright sky Dead Grass	Cold
		Bright sky Green grass Water	Fresh Bright Cool Open
		Dark sky Green grass	Dark Open

Five images together with the results of segmentation, object recognition and impression words estimation are shown in Table 3. For example, image at the first row is segmented into 3 regions, where each region is recognized as 'bright sky', 'bright sky' and 'green grasses'. Elimination the same recognized objects produces only 'bright sky' and 'green grass' as the result of object recognition. Then, each of 8 impression words is investigated by comparing the threshold value of each impression words with the size of the object, producing 'cool' and 'open' as the result of impression words estimation.

Images are stored into database together with objects and impression words contents. In retrieving stage, system do simple searching based on objects and/or impression words as retrieval keywords, and then displays the result images to user. Table 4 shows some query examples and three image results for each query.

Table 4. Examples of image retrieval.

Query	The results images
Cool and no Land and no Building	
	
	
Fresh and no Water	
	
	

#### 4. Conclusion

In this research, an image retrieval system based on object extraction and human *Kansei*, which enables retrieving images by using object names and/or human *Kansei* as retrieval keywords, has been developed and tested. Extracting objects is carried out automatically, and then, the recognized objects are used to estimate human *Kansei* by exploit the relationship between characteristics of objects in image and impression words representing human *Kansei*. The relationship between characteristics of objects and impression words is analyzed from questionnaires. Whereas most of *Kansei*-based image retrieval systems based primarily on characteristics of a whole image to produce impression words, this system derives impression words from characteristics of the extracted objects in images.

As a further study, more precise and widely applicable to describe relationship between wider variety of impression words and characteristics of objects must be required, while improving the segmentation method and developing human expression estimation is also important in order to estimate impression words for image consisted of wide area of human face.

#### References

- [1] C. Carson, S. Belongie, H. Greenspan, and J. Malik, "Region based Image Querying," Proc. 1997 IEEE Workshop on Content-Based Accesses of Image and Video Libraries, pp. 42-49, 1997.
- [2] Y. Li and L. G. Shapiro, "Consistent Line Clusters for Building Recognition in CBIR," Proc. 16th International Conference on Pattern Recognition (ICPR'02), vol. 3, pp. 952-957, 2002.
- [3] P. Duygulu and M. Bastan, "Translating Images to Words for Recognizing Objects in Large Image and Video Collections," Towards Category-Level Object Recognition, Lecture Notes in Computer Science Series vol. 4170, Springer Verlag, 2006.
- [4] S. Tanaka, M. Inoue, M. Ishiwaka, S. Inoue, "A method for Extracting and Analyzing *Kansei* Factors from Pictures," Proc. IEEE First Workshop on Multimedia Signal Processing, pp. 251-256, 1997.
- [5] T. Shibata, T. Kato, "*Kansei* Image Retrieval System for Street Landscape Discrimination and Graphical Parameters based on Correlation of Two Images," Proc. 1999 IEEE International Conference on Systems, Man, and Cybernetics, vol. 6, pp. 247-252, 1999.
- [6] K. Kuroda, M. Hagiwara, "An Image Retrieval System by Impression Words and Specific Object names-IRIS," Neurocomputing, vol. 43, pp. 259-276, 2002.
- [7] Y. Deng, B. S. Manjunath, "Unsupervised segmentation of color-texture regions in images and video," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 23, no. 8, pp. 800-810, 2001.
- [8] B. T. Luke, "K-Means Clustering," <http://fconyx.ncifcrf.gov/~lukeb/kmeans.html>