

# Automatic Classification for Various Images Collections Using Two Stages Clustering Method

Wan Hyun Cho<sup>1</sup>, In Seop Na<sup>2</sup>, Jun Yong Choi<sup>1</sup>, Tae Hoon Lee<sup>1</sup>

<sup>1</sup>Department of Statistics, Chonnam National University, Gwangju, Korea

<sup>2</sup>School of School of Electronic & Computer Engineering, Chonnam National University, Gwangju, Korea

Email: [whcho@jnu.ac.kr](mailto:whcho@jnu.ac.kr), [ypencil@hanmail.net](mailto:ypencil@hanmail.net), [abyss1225@hanmail.net](mailto:abyss1225@hanmail.net), [lth10916@hanmail.net](mailto:lth10916@hanmail.net)

Received 2012

## ABSTRACT

In this paper, we propose an automatic classification for various images collections using two stage clustering method. Here, we have used global and local image features. First, we review about various types of feature vector that is suitable to represent local and global properties of images, and similarity measures that can be represented an affinity between these images. Second, we consider a clustering method for image collection. Here, we first build a coarser clustering by partitioning various images into several clusters using the flexible Mean shift algorithm and K-mean clustering algorithm. Second, we construct dense clustering of images collection by optimizing a Gaussian Dirichlet process mixture model taking initial clusters as given coarser clustering. Finally, we have conducted the comparative experiments between our method and existing methods on various images datasets. Our approach has significant advantage over existing techniques. Besides integrating temporal and image content information, our approach can cluster automatically photographs without some assumption about number of clusters or requiring a priori information about initial clusters and it can also generalize better to different image collections.

**Keywords:** Automatic Classification; Images Collections; Clustering; Mean Shift; Gaussian Dirichlet Processing Mixture Model

## 1. Introduction

The general goal in the image clustering is to classify the different image objects or patterns in such a way that samples of the same cluster are more similar to one another samples belonging to different clusters. However, clustering task is a difficult problem due to many assumptions, different contexts and the variety of input data [1]. In the last few decades, it has been growing interests in developing effective and fast methods for classify an input image into different clusters. These methods are mainly divided in two types of clustering algorithms, such as supervised and unsupervised methods. In the supervised image clustering algorithms, the researchers incorporate a priori knowledge, such as the number of image clusters. Huang et al. [2] propose a hierarchical classification tree that is generated via supervised learning, using a training set of images with known class labels. The tree is next used to categorize new images entered into the database. Carson et al. [3] used a naive Bayes algorithm to learn image categories in a supervised learning scheme. The images are represented by a set of homogeneous regions in color and texture feature space, based on the “Blob-world” image representation. Yang et al. [4] propose a new clustering algorithm, re-

ferred to local discriminate models and global integration (LDMGI), which utilizes both manifold information and discriminate information for data clustering. They theoretically prove that K-means and DisKmeans are both special cases of LDMGI. They also show that LDMGI is a type of special clustering algorithm. Thus, they provide a new perspective to discover and understand the relationships between K-means (or DisKmeans) and other spectral clustering algorithms. Sleit et al. [5] propose Content Based Image retrieval (CBIR) scheme that extracts color, texture, and shape feature of images. Then, they group similar images together using K-mean clustering. They use the color histogram, Gabor filters, and Fourier descriptors for color, texture, and shape features respectively. The main restriction in supervised image clustering is that human intervention is required.

On the other hand, unsupervised methods aim at providing the correct number of image clusters without any a priori information. Goldberger et al. [6] combine discrete and continuous image models with information-theoretic based criteria for unsupervised hierarchical image-set clustering. The continuous image modeling is based on mixture of Gaussian densities. The unsupervised image-set clustering is conduct by the information

bottleneck principle. Krinidis et al. [7] present a supervised image clustering approach based on the image histogram, which is processed by the empirical mode decomposition (EMD). The Ensemble Empirical Mode Decomposition (EEMD), which provides noise resistance and assistance to data analysis, decomposes the image histogram into a number of Intrinsic Mode Functions (IMFs). The local maxima of the IMFs summation provide the desire number of image clusters and a combination of them is used as a criterion for image clustering.

In this paper, we present an unsupervised clustering method for large images dataset using two statistical clustering methods based on local and global invariant features. First, we think about a various types of feature vector that is suitable to represent local and global properties of images, and similarity measures that can be represented an affinity between these images. Next, we consider a clustering method for image collection. Here, we first build a coarser clustering by partitioning various images into several clusters using the Mean shift clustering and K-means clustering algorithms. Second, we construct dense clustering of images collection by optimizing a Gaussian Dirichlet process mixture model taking initial clusters as the derived coarser clustering. Finally, we have conducted the comparative experiments between our method and existing methods on various images datasets.

## 2. Extraction of Global and Local Feature

### 2.1. Global Image Feature

First, we consider global properties of color image. Here, we will use three kinds of feature information for clustering of given images. These are color feature, texture feature and shape feature. Here, we first consider color feature as color histograms [8]. It indicates the frequency of occurrences of every color in an image, and can be defined as a mass function. Our work is based on the HSV color histogram feature extraction. Second, we will consider texture feature as Discrete Wavelet Transform (DWT) [9]. Texture refers to visual patterns with properties of homogeneity that do not result from the presence of only a single color such as clouds and water. Texture features typically consist of contrast, uniformity, coarseness, and density. There are two basic classes of texture descriptors, namely, statistical model-based and transform-based. The former one explores the gray-level spatial dependence of textures and then extractes some statistical feature as texture representation. The latter approach is based on some transform such as DWT. Third; we consider a shape feature vector as moment invariants of image. Moment invariants have been frequently used as features for shape representation of object. They are

computed based on the information provided by both the shape boundary and its interior region.

### 2.2. Similarity Estimation

In order to verify the ability of distinguish between global feature vectors we compute the similarity or dissimilarity measures based on exponential the cosine distance for these two vectors. This measure between two feature vector  $\mathbf{f}_i$  and  $\mathbf{f}_j$  is defined as

$$S_c(i, j) = \exp\left(\frac{\langle \mathbf{f}_i, \mathbf{f}_j \rangle}{\|\mathbf{f}_i\| \|\mathbf{f}_j\|} - 1\right), \quad 1 \leq i, j \leq N.$$

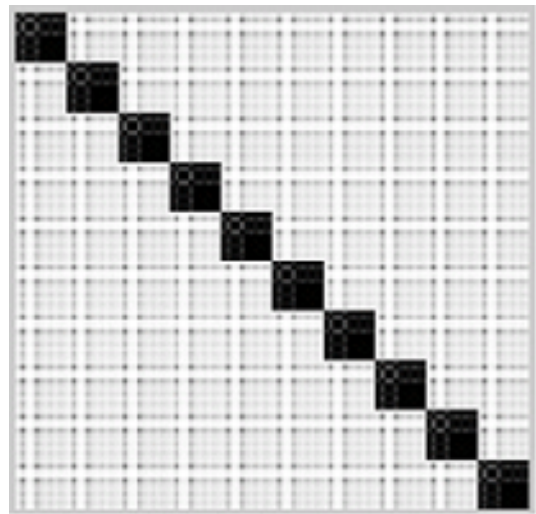


Figure 1. Similarity matrix between global feature vectors extracted for 110 images.

Figure 1 shows the similarity matrix between color feature vectors for 10 group images consisting of 11 images with same colors. In Figure 1, the main diagonal areas represent the similarities of images with the same color, but the non-main diagonal areas represent the similarities of images with different colors. Therefore, the images with same colors represent by pure black color, otherwise they represent by white color.

### 2.3. Local Image Feature

Here, we introduce image representation using the bag of visual words models based on local features. We have first various detectors and descriptors describing the image characters that are locally invariant with image rotation, scale transformation and illumination changes. We then describe a local features histogram that is made from bag of visual words using numerous local descriptors. A salient region in an image is a connected part of an image showing a significant and interesting image property. It is usually determined by the application of a region of interest detector to the image. If a region de-

tor returns only an exact position within the image, we also refer to it as interest point detector. The most important information that ideal region detectors give to us is the location of features, but other characteristics such as shape (scale) and orientation of a region of interest have to deliver additionally. Next, we have also discussed the interest feature descriptors and their characteristics. A descriptor is a process that takes information of features and image to produce descriptive information i.e. features' description, which are usually presented in form of features vectors. The descriptions then are used to match a feature to one in another image. And two important aspects that a descriptor has to satisfy are discriminative and invariant.

## 2.4. Local histogram of visual words

For image classification, we use the Bag-of-Visual Words approach, where images are represented as a histogram of visual words. The visual words denote local features extracted from the images and the vocabulary is learnt task-specifically from a training database [10-11]. The construction procedure of histogram features of visual words given from images goes as follows. First, we extract local feature descriptors from image patches around feature detectors which are invariant with scale or rotation change, and apply PCA transformation with these descriptors to reduce their dimensionality. Second, to create efficient codebook of visual words, we have partitioned the local descriptor space into several information regions using various clustering methods such as K-means clustering or GMM clustering model. Here, we can create a bag of visual words by choosing the center of each cluster as the visual word. Third, a bag of visual words is used as the codebook to build an image histogram of local features

## 3. Automatic Images Clustering

We automatically categories cluster images collections using two-stage clustering method.

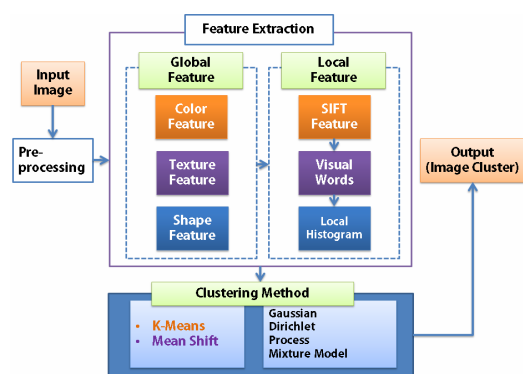


Figure 2. Block diagram of automatic images clustering

The first step builds initial coarser clustering by considering the contents of color images using Mean shift clustering and K-means clustering algorithm. The second step constructs accurate dense clustering by considering simultaneously the global and local features of color images using a Gaussian Dirichlet process mixture model. Fig. 2 shows the block diagram of two-stage clustering method using global and local features.

### 3.1. Coarser Clustering

First, we briefly review the Mean Shift clustering and K-means clustering algorithms. And then we built initial coarser clustering initial coarser clustering by considering the contents of color images using Mean shift clustering and K-means clustering algorithm. First, we briefly review the traditional Mean shift clustering procedure [12]. This is guaranteed to converge to a point where the gradient of density function is zero. Here, the key point of Mean shift clustering procedure is how to take a size of bandwidth. We have used the flexible bandwidth size obtained through multiple iterations of the implementation. Second, we briefly review the K-means clustering algorithm. The most common algorithm uses an iterative refinement technique. The algorithm is deemed to have converged when the assignments no longer change.

Here, we construct initial coarser clustering for experimental image collections using the Mean shift clustering algorithm with the flexible bandwidth size, and K-means clustering algorithm based on global and local feature vectors respectively. Here, global feature consist of color, texture and shape features. They are the histogram of RGB color space, histogram of coefficients generated by 2-step discrete wavelet transformation and invariant moments. Moreover, local feature is used by the histogram of visual words occurrences derived from applying K-means with SIFT descriptors.

### 3.2. Dense Clustering

Here, we first introduce a statistical theory of the Gaussian Dirichlet process mixture model. And then we construct accurate dense clustering by considering simultaneously the global and local features of color images using a Gaussian Dirichlet process mixture model. The Dirichlet process, denoted as  $\mathcal{G}$ , is a random measure on measures and is parameterized by the innovation parameter  $\alpha$  and a base distribution  $G_0$  [13]. One of the most important applications of the Dirichlet processes is as a nonparametric prior distribution of a mixture model. We want to model this data by means of nonparametric Bayesian formulation of the Gaussian Dirichlet process mixture model. For this purpose, since the number of mixture component is unknown, we have to consider the mixture model with countable infinite components.

Therefore, we will use a Dirichlet process mixture model as the prior distribution over the number of components generating the data, and we also assume the probability distribution of observations as the multivariate Gaussian distribution. Here, we assume a coarser clustering model of given image-sets obtained by Mean shift clustering or K-means algorithm as an initial clustering model for Gaussian Dirichlet process mixture model. If we apply Variational Bayesian inference principle for the Gaussian Dirichlet process mixture, then we can obtain the approximating likelihoods and posterior distributions  $q(\cdot)$  for all model parameters and latent clustering variables. They are represented with the following formulas. First, for the posterior distributions over the DP parameters, we have

$$q(v_k) = \text{Beta}(\beta_{k,1}, \beta_{k,2})$$

where

$$\beta_{k,1} = 1 + \sum_{n=1}^N q(z_n = k)$$

$$\beta_{k,2} = \alpha + \sum_{l=k+1}^K \sum_{n=1}^N q(z_n = l)$$

and

$$q(\alpha) = \text{Gamma}(\alpha | \hat{\eta}_1, \hat{\eta}_2)$$

where

$$\hat{\eta}_1 = \eta_1 + K - 1$$

$$\hat{\eta}_2 = \eta_2 - \sum_{k=1}^{K-1} [\psi(\beta_{k,2}) - \psi(\beta_{k,1} + \beta_{k,2})]$$

and  $\psi(\cdot)$  denotes the digamma function.

Second, regarding the posteriors over the likelihood parameters, we have

$$q(\Theta_k) = \mathcal{N}(\mu_k, \Lambda_k | \hat{\lambda}_k, \hat{\mathbf{m}}_k, \hat{\omega}_k, \hat{\Psi}_k)$$

where

$$\hat{\lambda}_k = \lambda_k + N_k$$

$$\hat{\mathbf{m}}_k = \frac{1}{\hat{\lambda}_k} (\lambda_k \mathbf{m}_k + N_k \bar{\mathbf{y}}_k)$$

$$\hat{\omega}_k = \omega_k + N_k,$$

$$\hat{\Psi}_k = \Psi_k + N_k \mathbf{S}_k + \frac{\lambda_k N_k}{\lambda_k + N_k} (\mathbf{y}_n - \bar{\mathbf{y}}_k)(\mathbf{y}_n - \bar{\mathbf{y}}_k)^T.$$

and we use the notation defined as

$$N_k = \sum_{n=1}^N q(z_n = k)$$

$$\bar{\mathbf{y}}_k = \frac{1}{N_k} \sum_{n=1}^N q(z_n = k) \mathbf{y}_n$$

$$\mathbf{S}_k = \frac{1}{N_k} \sum_{n=1}^N q(z_n = k) (\mathbf{y}_n - \bar{\mathbf{y}}_k)(\mathbf{y}_n - \bar{\mathbf{y}}_k)^T$$

Finally, the posteriors over latent clustering variables generating the clustering model is given as

$$q(z_n = k) \propto \hat{\pi}_k(\mathbf{v}) \hat{p}(\mathbf{y}_n | \Theta_k)$$

where

$$\hat{\pi}_k(\mathbf{v}) = \exp\{\sum_{l=1}^{k-1} \langle \log(1 - v_l) \rangle + \langle \log v_k \rangle\}$$

and

$$\hat{p}(\mathbf{y}_n | \Theta_k) = \exp\left\{-\frac{d}{2} \log 2\pi + \frac{1}{2} \langle \log |\Lambda_k| \rangle - \frac{1}{2} \langle (\mathbf{y}_n - \mu_k)^T \Lambda_k (\mathbf{y}_n - \mu_k) \rangle\right\}$$

As a last step, after conducting the updates of all posterior distributions and likelihood parameters at each iteration of the variational Bayesian inference algorithm for the Gaussian Dirichlet process mixture, the estimates  $Z$  of the latent clustering variables must be also update. Then, we have obtained the cluster membership of each image by maximization of posterior distribution over  $k$ . Hence, if each image is assigned to the cluster with cluster membership

$$\hat{z}_n = \text{argmax}_{1 \leq k \leq K} q(z_n = k),$$

then we obtained final clustering result for given image datasets.

## 4. Experimental results

### 4.1. Synthetic Image Data

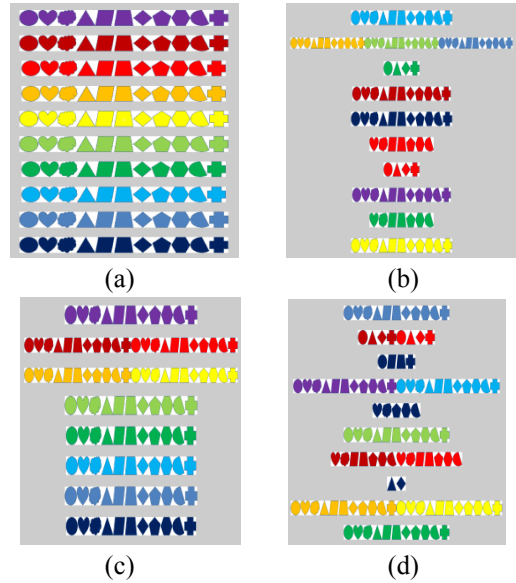


Figure 3. Coarser clustering results using means-shift and K-means based on color histogram with 8 bins and 4 bins for each channel: (a) Mean shift(8 bins), (b) K-means(8 bins), (c) Mean shift(4 bins), (d) K-means(4 bins)



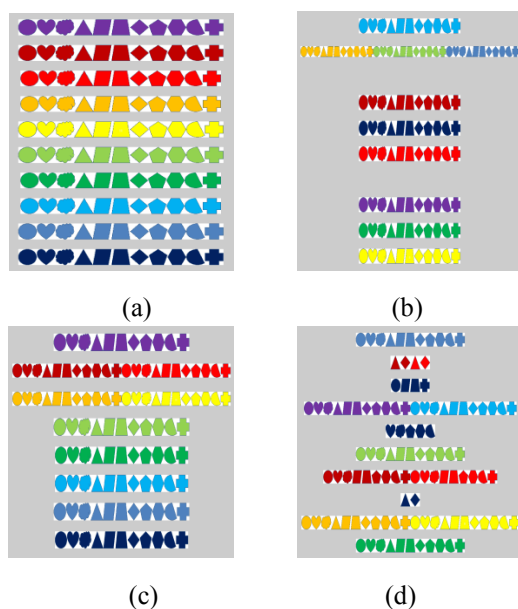


Figure 4. Dense clustering results using means-shift + VDPGMM and K-means+VDPGMM based on color histogram with 8 bins for each channel: (a) Mean shift + VDPGMM(8 bins), (b) K-means + VDPGMM(8 bins), (c) Mean shift+VDPGMM(4 bins), (d) K-means + VDPGMM

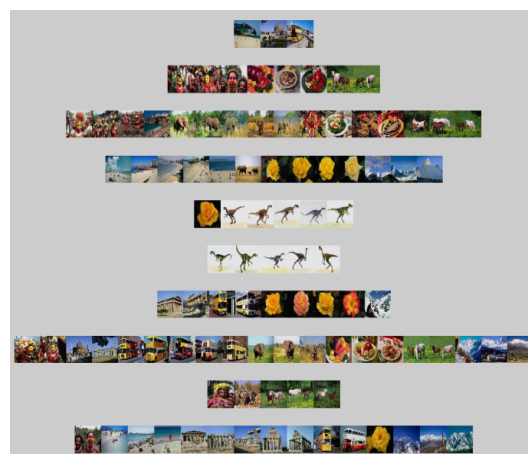
## 4.2. Real image data

We conducted experiments using Matlab on an image database that is collected from internet public data set CBIR. Our image database consist of 10 different groups, namely, Festival, Bus, Horse, Rose, Beach, Dinosaurs, Food, Scene, Elephant, The Parthenon temple as shown in Fig. 5. Each group contains 10 similar images. The spatial resolution of each image is size of  $128 \times 128$  pixels. We first test our clustering algorithm using global features such as color, texture, shape features, K-means and GDPM clustering method. In the coarser clustering step, we build initial clusters by considering the global features of images and K-means clustering method. And then in the dense clustering step, we construct fine clusters using a Gaussian Dirichlet process mixture model taking an initial clustering model as K-means clustering results. Our experimental results show that we can cluster properly an example image collection into 10 groups. Our clustering result is shown in Fig 5. Second, we test our clustering algorithm using local features generated by SIFT, visual word and local image histogram, K-means and GDPM clustering method. In the coarser clustering step, we build initial clusters by considering the local features of images and K-means clustering method. And then in the dense clustering step, we construct fine clusters using a Gaussian Dirichlet process mixture model taking an initial clustering model as K-means clustering results. Our experimental results show that we can cluster properly an example image collection into 10 groups.

Our clustering result is shown in Fig 5.

Table 1. Comparative results for global and local feature.

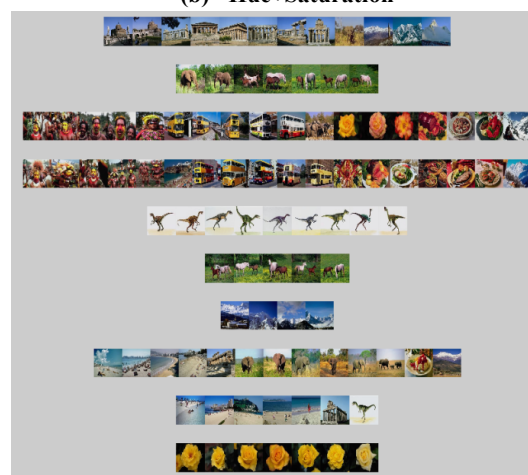
| Feature | Precision | Recall | F-measure |
|---------|-----------|--------|-----------|
| Global  | 64.0      | 54.0   | 55.6      |
| Local   | 50.1      | 49.0   | 43.8      |



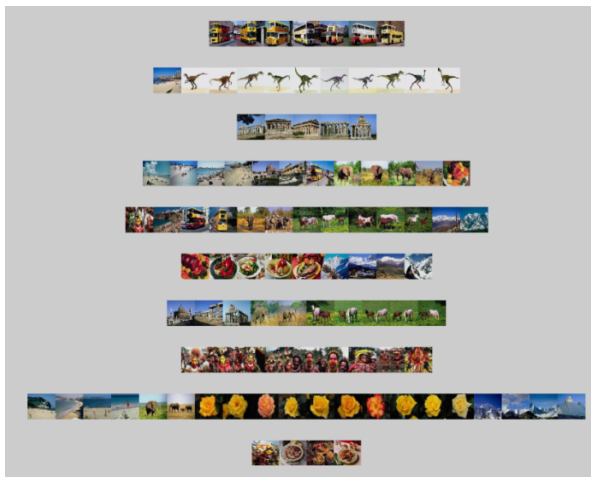
(a) DWT



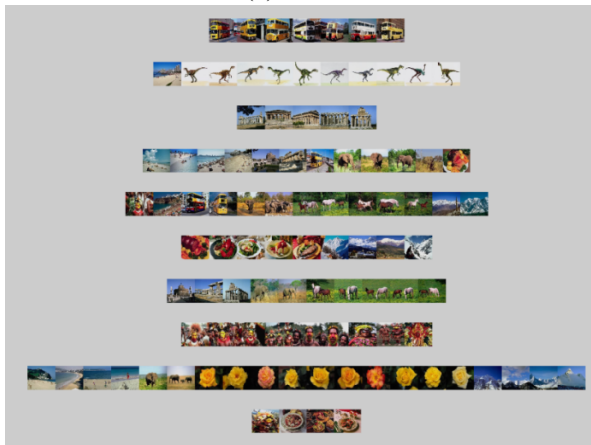
(b) Hue+Saturation



(c) Hue+Saturation+DWT



(d) KMeans



(e) KMeans+GDPM

Figure 5. Dense and coarser clustering results

## 5. Conclusion

In this paper, we present methods for clustering images using Mean Shift algorithm and Gaussian Dirichlet mixture model. Our approach has significant advantage over existing techniques. Besides integrating temporal and image content information, our approach can cluster automatically photographs without some assumption about number of clusters or requiring a priori information about initial clusters and it can also generalize better to different image collections.

## 6. Acknowledgements

This work was supported in part by the Korea Research Foundation Grant by the Korean government (KRF 2012-0002487). And "This research was supported by the MKE(The Ministry of Knowledge Economy), Korea, under the 3D and Smart TV Competitiveness Program support program (NIPA-2012-I2901-12-1031) supervised

by the NIPA(National IT Industry Promotion Agency)

## REFERENCES

- [1] K. Barnard, P. Duygulu, and D. Forsyth, Clustering art, Present in Comput. Vis. Pattern Recognition, Dec., 2001, pp 434-441.
- [2] J. Huang, S. R. Kumar, and R. Zabith, An automatic hierarchical image classification scheme, in ACM Conf. Multimedia, Set., 1998, pp 219-228.
- [3] C. Carson, S. Belongie, H. Greenspan, and J. Malik, Blobworld: Image segmentation using Expectation-Maximization and Its Application to Image Querying, IEEE Tran. On PAMI, Vol. 24, No. 8, pp 1026-1038, 2002.
- [4] Y. Yang, D. Xu, F. Nie, S. Yan, and Y. Zhung, Image Clustering Using Local Discriminant Models and Global Integration, IEEE Tran. On Image Processing, Vol. 19, No. 10, pp 2761-2773, 2010.
- [5] A. Sleit, A. Abugalhoun, M. Qatawneh, M. Ai-Sharief, R. Al-Jabaly and Ola Karajeh, Image clustering using Color, Texture and Shape features, KSII Tran. On Inter. And Infor. Syst., Vol. 5, No. 1, pp 212-227, 2011.
- [6] J. Goldberger, S. Gordon, and H. Greenspan, Unsupervised Image-Set Clustering Using an Information Theoretic Framework, IEEE Transaction on Image Processing, Vol. 15, 2006, pp 449-458.
- [7] S. Krinidis, M. Krinidis, and V. Chatzis, An Unsupervised Image Clustering Method Based on EEMD Image Histogram, Jour. Of Infor. Hiding and Multi. Sig. Processing, Vol. 3, No. 2, pp 152-163, 2012.
- [8] N. Sharma, S. P. Rawat, S and J. Singh, S, Efficient CBIR using Color Histogram Processing, Signal & Image Processing : An International Journal, Vol. 2, No.1, pp 94-112, 20011.
- [9] K. H. Ghazali, M. F. Mansor, M. M. Mustafa and A. Hussain, Feature Extraction Technique using Discrete Wavelet Transform for Image Classification, The 5th Student conference on Research and Development, 11-12 December 2007, Malaysia.
- [10] R. J. Lopez-Sastre, T. Tuytelaars, F. J. Acevedo-Rodriguez, and S. Maldonado-Bascon, Towards a more discriminative and semantic visual vocabulary. Computer Vision and Image Understanding, Vol. 115, pp 415-425, 2011.
- [11] T. Deselaers, L. Pimenidis, and H. Ney, Bag-of-Visual Words Models for Adult Image Classification and Filtering, 208 IEEE
- [12] C. Xiao and M. Liu, Efficient Mean-shift Clustering Gaussian KD-Tree, Pacific Graphics, Vol. 9, No. 7, pp , 2010
- [13] D. M. Blei and M. I. Jordan, "Variational Inference for Dirichlet Process Mixtures," Bayesian Analysis, vol. 1, no. 1 pp 121-144, 2006.