Unsupervised Content-Based Image Retrieval based on Mean Shift Clustering

Saman Bashbaghi[1], Mostafa Parchami[1], Hassan khotanlou[1],

[1] Computer Engineering Department, Bu-Ali Sina Uinversity, Hamedan, Iran
{bashbaghi, parchami, khotanlou}@basu.ac.ir

***Abstract*—According to increase in the number of digital images, searching on them becomes more important task to everyone. In this paper, we propose a new method for content-based image retrieval. Our method is based on automatic segmentation and indexing on heterogeneous and unlabeled images and is tolerant with different conditions such as any rotation and lighting conditions. First of all, for training phase the image is converted to YCbCr color space format, and then is segmented by the Mean Shift clustering algorithm. For each segment, some features are extracted from the original image in HSI format. Both region-based and block-based feature extraction is used. Then features are clustered into K categories by a K-means algorithm and a 1-NN classifier is used to create the indexed database in order to retrieve similar images. For evaluating the method, ZuBuD database has been used and 96.43% recognition rate of the first retrieved image was obtained.**

***Keywords* — Content-Based Image Retrieval, Image Segmentation, Mean Shift Clustering, K-means Clustering and 1-NN Classifier.**

I. INTRODUCTION

A huge amount of images are generated in our everyday life as the fast growth of advanced digital capturing devices for multimedia, such as digital camera and mobile photography phone. Through WWW, the collective image repository will be further bigger and bigger because of the speeding exchange of these life images. With this huge distributed and heterogeneous information, people want to search and make use of images over their contents. A great challenge is finding out accurate ways of searching images. As a result, how to access the growing heterogeneous repositories effectively and efficiently has been becoming an attractive research topic for multimedia processing. For this purpose, many general purpose image retrieval systems have been developed. Basically, images can be retrieved in two ways: first, text based and second, content-based or query by example based [1]. The term CBIR1 seems to have originated in the earlier 90's [1]. CBIR includes research on: automatic feature extraction [4], [5], automatic feature extraction with a semantic content [1] and data representation [6], [7]. Low-level image feature extraction is the basis of CBIR systems. CBIR techniques use features such as texture, color and shape to represent images and retrieves images relevant to the query image from the image database. Among those image features, texture features has been shown very effective and subjective [1]. To perform CBIR, image features can be either extracted from the entire image or from regions. As it has been found that users are usually more interested in specific regions rather than the entire image, most current CBIR systems are region-based. Global feature based retrieval is comparatively simpler.

Representation of images at region level is proved to be more close to human perception system [9]. Segmentation is very important to image retrieval. Both the shape feature and the layout feature depend on good segmentation. Automatic image segmentation is a difficult task. A variety of techniques have been proposed in the past, such as curve evolution [10], energy diffusion [11], and graph partitioning [12]. Many existing segmentation techniques work well for images that contain only homogeneous color regions, such as direct clustering methods in color space [13].

Some systems design their own segmentations in order to obtain the desired region features during segmentation, be it color, texture, or both [14], [15], [16], [17], [18], [19], [20]. These algorithms are usually based on k-means clustering of pixel/block features. In [16], an image is firstly segmented into small blocks of size 4∗4 from which color and texture feature are extracted. Then k-means clustering is applied to cluster the feature vectors into several classes with each class corresponding to one region. Blocks in same class are classified into same region. A KMCC (k-means with connectivity constraint) is proposed in [20] to segment objects from images. It is extended from the K-means algorithm. In this algorithm, the spatial proximity of each region is taken into account by defining a new center for the k-means algorithm and by integrating the k-means with a component labeling procedure.

According to increase of the number of images in different subjects and different conditions in our world and since searching and retrieving these images is important to everyone who needs to work with them, the problem of content-based image retrieval is a significant problem in the image processing and information retrieval scope. As we know labeling images is hard and expensive work to human and corporations. Also sometimes the label may be inappropriate and the point is usually the labels do not describe images as well as we interpret them by natural features. So we concentrate on designing an unsupervised CBIR system that works with unlabeled images to reduce the cost of labeling process and to improve consistency of our system with any type of images. In this research, we proposed a new CBIR system based on image segmentation. An unsupervised learning is performed to adapt our system with any kind of unlabeled images which have any kind of condition. Most of previous CBIR systems work well only with some special conditions, but because of good segmentation and appropriate feature extraction as image description, our experiments show that it will work for many kinds of image databases.

[1] Content-Based Image Retrieval

This paper structure consists of proposed method, experimental results, conclusion and future works.

II. PROPOSED METHOD

Proposed method in this research is based on image segmentation with Mean Shift clustering. In this method, features are extracted from segments (Region-based feature extraction) and also entire image (block-based feature extraction). This method consists of two major stages: training stage and retrieving stage. In this section, we describe each parts of our proposed method in details.

A. *Training Stage*

Fig. 1 shows the block diagram of the training stage which is based on work in [23]. This stage consists of two phases as shown in Fig. 1 and used to create the indexed image database for the next stage (Retrieval).

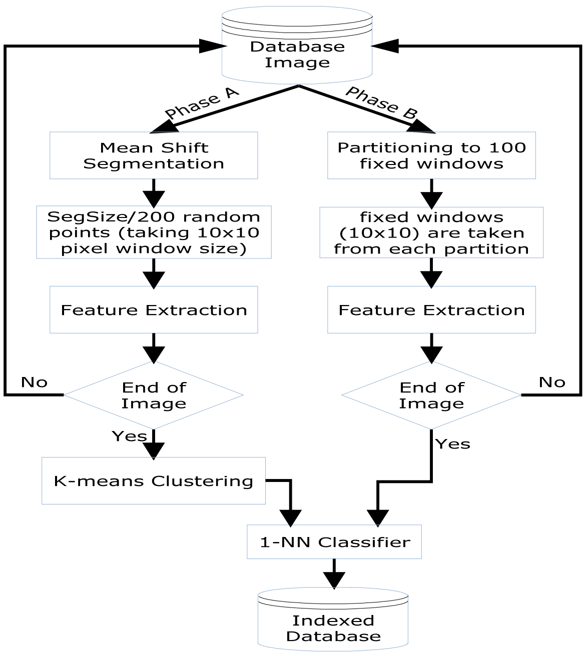


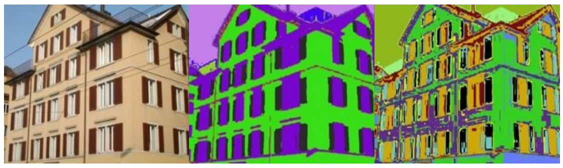
Figure 1. Block Diagram of Training Stage.

1) *Phase A*

During the first phase, all images in the database, which are in RGB color space, are converted to YCbCr color space. Each image is then segmented by Mean Shift clustering. The mean shift algorithm is a nonparametric clustering technique which does not require prior knowledge of the number of clusters, and does not constrain the shape of the clusters [24].

In contrast to the classic K-means clustering approach [25], there are no embedded assumptions on the shape of the distribution or the number of modes/clusters. Mean Shift was first proposed by Fukunaga and Hostetler [26], later adapted by Cheng [27] for the purpose of image analysis and more recently extended by Comaniciu, Meer and Ramesh to low-level vision problems, including, segmentation [28], adaptive smoothing [28] and tracking [29]. Most of previous works, which are based on image segmentation, used K-means clustering for segmentation.

In our work, we use Mean Shift instead of K-means. In Fig. 2 We compared these algorithms with the same number of clusters.

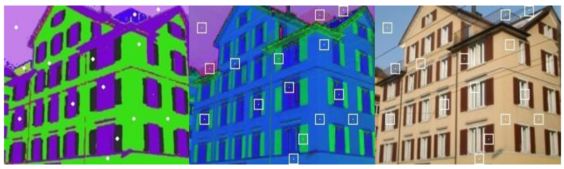


(a) (b) (c)

Figure 2. Comparing Mean Shift and K-means clustering for image segmentation. (a) Original Image. (b) Mean Shift segmentation. (c) K-means segmentation.

As shown in Fig. 2. Mean Shift clustering (Fig. 2 b) segmented the original image (Fig. 2 b) better than K-means clustering (Fig. 2 c).

After the image is segmented, we extract features from these segments. For this purpose, we pick some points randomly from each segment. As a part of preprocessing phase, we convert the original image into HSI format. After taking each of the random points of each segment as the center of a window (Fig. 3 a), we then open a squared window of size 10×10 pixels around it (Fig. 3 b) and then the features are extracted from original image in HSI format (Fig. 3 c).



(a) (b) (c)

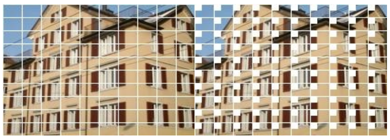
Figure 3. Picking random points from each segment. (a) Random points in each cluster. (b) Opened windows around each point. (c) Features taken from original image in HIS format.

The size of each segment is divided by 200 because with these opened windows we can cover 50% of segment size for analyzing and extracting features. The following features are extracted for each window: the mean, the standard deviation and the homogeneity obtained from the co-occurrence matrix in three channels: hue (H), saturation (S) and Intension (I) of HSI format.

After extracting features for each segment, at the end of this phase, we cluster these features into K clusters by K-means algorithm. We can select K, according to the number of dataset's categories. It's necessary to say our method is rotation invariant because of segmentation and selected features.

2) *Phase B*

During this phase, as shown in Fig 4. Each image in the database is uniformly partitioned into 100 regions (Fig. 4 a). Then we take a window of 10×10 pixels for each of these 100 subimages (Fig. 4 b), then the same feature described before are extracted for each window. Features were: the mean, the standard deviation and the homogeneity obtained from the co-occurrence matrix in three HSI channels.



(a) (b)

Figure 4. Phase B of training stage.

3) *Creating Indexed Database*

In this section, to create the indexed database we measure the similarity between clustered features came from Phase A and features came from Phase B by presenting them to a 1-NN classifier. By this work we can measure the similarity of each window of Phase B with each segment from Phase A. At the end of this section, we construct the indexed database that contains the number of Phase B's windows that are classified in each cluster.

B. Retrieving Stage

In this stage an image is presented to the system as a query image. The features are then extracted from the query image as shown in the Phase B in Fig 1. As a result we get 100 feature vectors that describe the query image, and then a 1-NN classifier is used to classify each of these describing features in the classes have been created during training stage by K-means clustering. Finally a vector that shows the probability of each category contained in the query image is obtained. Euclidian distance is used to calculate distance between the resulted vector and vectors in the indexed database and then ten most similar images are retrieved as the final results.

III. EXPERIMENTAL RESULTS

In this section we present our experimental results to test our system and then we compare our system with other CBIR systems. For testing our approach we have used ZuBuD[2] database.

The Zurich Building Image Database (ZuBuD) was created by the Swiss Federal Institute of Technology in Zurich and is described in more detail in [30], [31]. The database consists of two parts, a training part of 1005 images of 201 buildings, 5 of each house. The images are of size 640×480 and were taken from different positions using different cameras under different lighting conditions. The query set consists of 115 images, which are not included in the database and acquired by another camera under different condition. In Fig. 5 Some samples of ZuBuD database are shown.

Figure 5. Some original images from ZuBuD database.

We presented each of 115 images to the system and retrieved 10 most similar images to the query image from indexed database. To measure effectiveness [3] of our system and evaluate its performance, we have used the following measures, Precision (P) and Recall (R) [32]:

$$P = \frac{\#\,Relevant\ images\ retrieved}{\#\,Images\ retrieved}\ x\,100 \qquad (1)$$

$$R = \frac{\#\,Relevant\ images\ retrieved}{\#\,Total\ relevant}\ x\,100 \qquad (2)$$

The first measure (1) represents the number of relevant images retrieved with respect to the total number of images asked to be retrieved. The second measure (2) represents the relevant images retrieved with respect to the total number of relevant images.

Fig. 6 shows comparison of average precision versus recall of our proposed method against proposed method in [23] on the ZuBuD database. As shown in Fig. 6 the precision of our proposed method at recall 0.2 is 97.7%, which means for the first relevant image retrieved from five relevant images in the database, the precision is 97.7%.

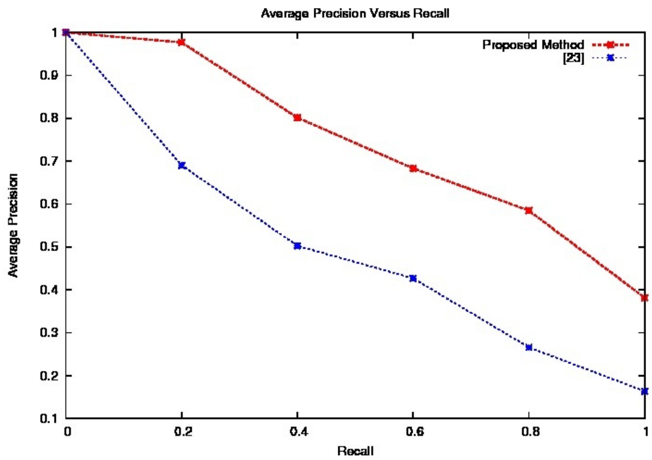


Figure 6. Performance of our method against method in [23].

As shown in Fig. 7 The classification recognition rate in the first retrieved image in [33], [34], [35] and [36] are 95.65%, 93%, 89.6% and 59.13% respectively. In our method the classification recognition rate on first retrieved image is 96.43%.

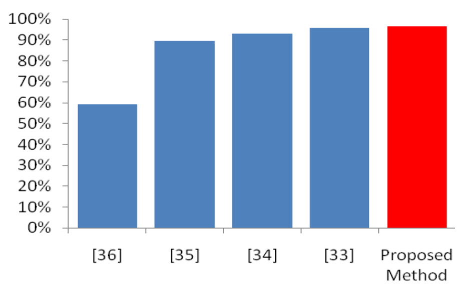


Figure 7. Comparing the Classification Recognition Rate

[2] http://www.vision.ee.ethz.ch/showroom/zubud/

[3] MOE

In Fig. 8 and Fig. 9 two retrieval examples are shown. As shown in Fig 8. first five retrieved images (Fig 8. b-f) are totally relevant to query image (Fig. 8 a), which means all five relevant images to this query are retrieved correctly.

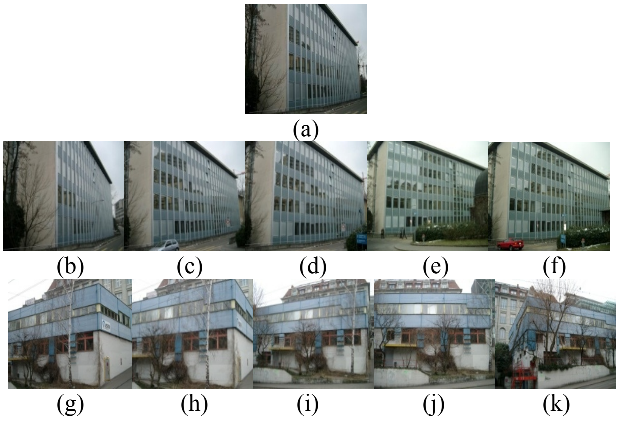


Figure 8. One retrieval example. (a) Query image, (b-f) First five retrieved images. (g-k) Other retrieved images.

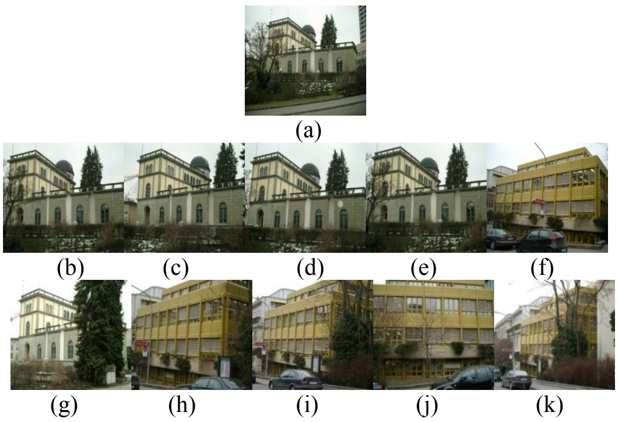


Figure 9. Another retrieval example. (a) Query image. (b-f) From the first five retrieved images, four images are relevant to the query image and fifth image is not relevant to the query image. (g-k) Other retrieved images. Sixth image is relevant to the query image.

IV. CONCLUSION AND FUTURE WORKS

As results shows, Mean Shift clustering has better performance than K-means clustering for segmentation. The proposed method is rotation invariant and is tolerant with any conditions because of selecting features from segmented objects. In this method, we don't also need to label images manually.

To improve the proposed CBIR system, we can use better features such as interest points and FCM[4] clustering can be used instead of K-means for describing features clustering. We are going to test our system on other databases.

REFERENCES

[1] Liu. Ying, Zhang. Dengsheng, Lu. Guojun, Ma. Wei-ying, A survey of content-based image retrieval with high level semantics, *Pattern Recognition - PR* , vol. 40, no. 1, pp. 262--282, 2007.

[2] J. Eakins, M. Graham, Content-based image retrieval, *Technical Report*, University of Northumbria at Newcastle, 1999.

[4] Fuzzy C-Means

[3] I.K. Sethi, I.L. Coman, Mining association rules between low-level image features and high-level concepts, *Proceedings of the SPIE Data Mining and Knowledge Discovery*, vol. III, pp. 279--290, 2001.

[4] A. Bimbo, A Perspective View on Visual Information Retrieval Systems, *Content Based Access of Image and Video Libraries*, Workshop on Vol. 21:108--109, IEEE 1998.

[5] R. Datta, D. Joshi, Jia. Li, J.Z. Wang, Image Retrieval: Ideas, Influences, and Trends of the New Age, *ACM Computing Surveys*, 40(2), April 2008.

[6] Y. Rui, Th.S. Huang, Sh.F. Chang, Image Retrieval: Currents techniques, Promising Directions, and open Issues. *Journal of Visual Communication and Image Representation 10*, pp. 39--62, 1999.

[7] J. Li, J.Z. Wang, Real-Time Computerized Annotation of Pictures. *Proc. Of the 14th annual ACM Int*. Conference on Multimedia, pp. 911--920, 2006.

[8] A. Mojsilovic, B. Rogowitz, Capturing image semantics with low-level descriptors, *Proceedings of the ICIP*, pp. 18--21, 2001.

[9] F. Jing, M. Li, L. Zhang, H.J. Zhang, B. Zhang, Learning in region-based image retrieval, *Proceedings of the International Conference on Image and Video Retrieval* (CIVR2003), pp. 206--215, 2003.

[10] H. Feng, D.A. Castanon, W.C. Karl, A curve evolution approach for image segmentation using adaptive flows, *Proceedings of the International Conference on Computer Vision (ICCV'01)*, pp. 494--499, 2001.

[11] W.Y. Ma, B.S. Majunath, Edge flow, a framework of boundary detection and image segmentation, *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 744--749, 1997.

[12] J. Shi, J. Malik, Normalized cuts and image segmentation, *IEEE Trans. Pattern Anal. Mach. Intell. (PAMI)* 22 (8), pp. 888--905, 2008.

[13] D. Comaniciu, P. Meer, Robust analysis of feature spaces: color image segmentation, *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 750--755, 1997

[14] P.L. Stanchev, D. Green, B. Dimitrov, High level color similarity retrieval, *Int. J. Inf. Theories Appl*. 10 (3), pp. 363--369, 2003.

[15] K.A. Hua, K. Vu, J.H. Oh, SamMatch, a flexible and efficient sampling-based image retrieval technique for large image databases, *Proceedings of the Seventh ACM International Multimedia Conference (ACM Multimedia'99)*, pp. 225--234, 1999.

[16] J.Z. Wang, J. Li, G. Wiederhold, SIMPLIcity, semantics-sensitive integrated matching for picture libraries, *IEEE Trans. Pattern Anal. Mach. Intell. 23* (9), pp. 947--963, 2003.

[17] C.P. Town, D. Sinclair, Content-based image retrieval using semantic visual categories, *Society for Manufacturing Engineers*, Technical Report MV01-211, 2001.

[18] J.R. Smith, C.S. Li, Decoding image semantics using composite region templates, *IEEEWorkshop on Content-Based Access of Image and Video Libraries (CBAIVL-98)*, pp. 9--13, June 1998.

[19] W.K. Leow, S.Y. Lai, Scale and orientation-invariant texture matching for image retrieval, in: M.K. Pietikainen (Ed.), Texture Analysis in Machine Vision, *World Scientific*, Singapore, 2000.

[20] V. Mezaris, I. Kompatsiaris, M.G. Strintzis, An ontology approach to object based image retrieval, *Proceedings of the ICIP*, vol. II, pp. 511--514, 2003.

[21] M. Lybanon, S. Lea, S. Himes, Segmentation of diverse image types using opening and closing, in *Proc. IEEE Int. Conf. on Image Proc.*, 1994.

[22] M. Hansen, W. Higgins, Watershed-driven relaxation labeling for image segmentation, in *Proc. IEEE Int. Conf. on Image Proc.*, 1994.

[23] Felix. Serrano, Carlos. Avilés, Sossa. Humberto, Juan. Villegas, Gustavo. Olague, Unsupervised Image Retrieval with Similar Lighting Conditions, Conference: *International Conference on Pattern Recognition - ICPR* , pp. 4368--4371, 2010.

[24] D. Comaniciu, P. Meer, Mean shift, A robust approach toward feature space analysis. *IEEE Trans. Pattern Anal. Machine Intell.*, 24:603--619, 2002.

[25] R. Duda, P. Hart , D. Stork: Pattern Classification, *Wiley*, 2001

[26] K. Fukunaga, L. Hostetler, The estimation of the gradient of a density function, with applications in pattern recognition. *IEEE Transactions on Information Theory*, 21(1), pp. 32--40, 1975.

[27] Y. Cheng, Mean shift, mode seeking, and clustering. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 17(8), 790--799, 1995.

[28] D. Comaniciu, P. Meer, Mean shift: A robust approach toward feature space analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(5), pp. 603--619, 2002.

[29] D. Comaniciu, V. Ramesh, P. Meer, Kernel-based object tracking. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 25(5), pp. 564--577, 2003.

[30] H. Shao, T. Svoboda, T. Tuytelaars, L.V. Gool, HPAT Indexing for Fast Object/Scene Recognition based on Local Appearance. In *Proc. CIVR*, LNCS 2728, pp. 71--80, Urbana-Champaign, IL, July 2003.

[31] H. Shao, T. Svoboda, L.V. Gool, ZuBuD – Zurich Buildings Database for Image Based Recognition. *Technical Report No. 260, Computer Vision Lab, Swiss Federal Institute of Technology*, Zurich, Switzerland, Apr. 2003.

[32] R. Baeza-Yates, B. Ribeiro-Neto, Modern Information Retrieval, *Addison Wesley*, 1999.

[33] Raphael. Maree, Pierre. Geurts, Louis. Wehenkel, Content-Based Image Retrieval by Indexing Random Subwindows with Randomized Trees, *Asian Conference on Computer Vision - ACCV* , pp. 611--620, 2007.

[34] T. Deselaers, D. Keysers, H. Ney, Classification Error Rate for quantitative Evaluation of Content-based Image Retrieval Systems, *Proc. ICPR*, pp. 505--508, 2004.

[35] S. Obdr zalek, J. Matas, Sub-linear Indexing for Large Scale Object Recognition, *Proc. BMVC*, pp. 1--10, 2005.

[36] H. Shao, T. Svoboda, V. Ferrari, T. Tuytelaars, V.L. Gool, Fast Indexing for Image Retrieval Based on Local Appearance with Re-ranking, *Proc. IEEE ICIP*, pp.737--749, 2003.