Blood Cell Image Retrieval System Using Color, Shape and Bag of Words

Mohammad Reza Zare[1] and Woo Chaw Seng[2]

[1]School of Information Technology, Monash University Malaysia
mohammad.reza@monash.edu
[2]Faculty of Computer Science and Information Technology, University of Malaya, Malaysia
cswoo@um.edu.my

Abstract. The ever increasing number of medical images in hospitals urges on the need for generic content based image retrieval systems. These systems are in an area of great importance to the healthcare providers. The first and foremost function in such system is feature extraction. In this paper, different feature extraction techniques have been utilized to represent medical blood cell images. They are categorized into two groups; low-level image representation such as color and shape analysis and local patch-based image representation such as Bag of Words (BoW). These features have been exploited for retrieving similar images. We have also used a generative model such as Probabilistic Latent Semantic Analysis (PLSA) on extracted BoW for retrieval task. Lastly, the retrieval results obtained from all the above features are integrated with one another to increase the retrieval performance. Experimental results using four different classes of 600 blood cell images showed 92.25% of retrieval accuracy.

Keywords: Image Retrieval. Feature Extraction. BoW. PLSA. CBIR.

1 Introduction

The growth of multimedia content production has been accelerated due to the wide availability of digital devices. Medical blood cell images are one of the most popular among the variety of multimedia content. The latest computer systems can store a very huge amount of medical images. Such medical image databases are the key component in diagnosis and preventive medicine. Therefore, there is an increased demand for a computerized system to manage these valuable resources. In addition, managing such data demands high accuracy since it deals with human life.

Traditional text based image retrieval is becoming infeasible due to the huge amount of blood cell images and it is also difficult to achieve a consistent diagnose during microscopic evaluation due to subjective impressions of observers. The solution to this problem is Content-Based Image Retrieval (CBIR); This would allow us to search based on the content rather than the keywords and meta-data descriptions. The CBIR performance strongly lays on the techniques used to represent images. There are two main approaches in CBIR: (i) those techniques that directly extract low level visual features from the images, and (ii) those techniques that represent images by

C.K. Loo et al. (Eds.): ICONIP 2014, Part III, LNCS 8836, pp. 218–225, 2014.
© Springer International Publishing Switzerland 2014

local descriptors. Low level visual features such as color, shape and texture have been used to classify and retrieve medical blood cell images in various studies [1-5].

Recently, more promising studies have been focused on local descriptors. SIFT feature is one of the most widely used local descriptor in object recognition tasks. With the advances of this local feature, researchers in the field of computer vision have attempted to resolve object classification problems by a new approach known as Bag of Words (BoW). In recent years, many studies have successfully exploited this feature in general scene and object recognition tasks [6-8] due to its simplicity, discrete representations and simple matching measures in preserving computational efficiency. The use of BoW model can also be found in medical image classification and retrieval tasks [9-12].The analysis on the results obtained from their studies proven that BoW performs better than other low level features. Andre *et al.* in [11] developed a system for endomicroscopy video retrieval. BoW has been employed in their study in order to produce visual and semantic outputs which are consistent with each other. In another medical image classification work, BoW was combined with other image representation techniques such as LBP, pixel value and Edge Histogram Descriptor with two different feature fusion schemes; Low Level and High Level [10]. The results obtained by this group clearly show that feature fusion methods outperform the results obtained by using a single feature in classification task. However, they have analyzed the results obtained by different feature extraction techniques and its proven that BoW features perform better than other feature representation used in that work.

BoW also works over probabilistic tools such as Latent Dirichlet Allocation (LDA) [18] and Probabilistic Latent Semantic Analysis (PLSA) [15], and has been successfully applied in image retrieval and annotation [13-14]. Although PLSA was originally proposed in the context of text document retrieval, it has also been applied to various computer vision problems such as classification and images retrieval where we have images as documents and the discovered topics are object categories (e.g. airplane, sky). Zare *et. al.* in [12] employed PLSA approach for classification of medical X-ray images in two different techniques; one to annotate medical X-ray images and the other one is to retrieve top five similar images to the query image. Classification accuracy rate obtained by this approach showed tremendous improvement compared to classification rate obtained by flat SVM classifier. Capturing meaningful aspects of images as well as generating low-dimensional and robust image representation can be considered as another ability of such tools which has been studied in various studies [14].

The purpose of this study is to improve the retrieval performance of the previous experiment [5] with a larger database. BoW has been employed as one of the feature extraction technique. PLSA based image retrieval has also been explored in this experiment. Then, intersections of the results obtained by this approach with the ones acquired by color and shape analysis are taken as final retrieval results.

The rest of the paper is organized as follows: Section 2 discusses the components of the proposed CBIR system for blood cell images. Experimental results and discussion are reported and analyzed in Section 3. Finally, the overall conclusion of this study is presented in Section 4.

2 Methodology

The proposed system is made up two modules; Feature Extraction and Image Retrieval, each has specific functionalities which will be described in detail below.

2.1 Feature Extraction Module

Feature extraction plays an important role in the performance of any image retrieval system because it can produce significant impact on the retrieval results. Different approaches for feature extraction have been employed in this experiment as explained below.

2.1.1 Color Analysis

Color is a fundamental characteristic of image content and it is a frequently used visual feature for CBIR. Color is a powerful descriptor that simplifies object recognition.

Histogram is a very commonly used color descriptor technique. Color histogram is obtained by quantizing the color space and counting the number of pixels that fall in each discrete color.

Color histogram is used to represent distribution of colors in an image in CBIR application. There are various techniques in measuring the dissimilarity between distributions of such features. In this research, Bhattacharya coefficient has been used to compare the color histograms between query image and images in the database. This is to indicate the relative closeness of the two sample images.

Calculating the Bhattacharya coefficient involves a rudimentary form of integration of the overlap of the two samples. The interval of the values of the two samples is split into a chosen number of partitions, and the number of members of each sample in each partition is used in the following formula:

$$\text{Bhattacharya} = \sum_{i=0}^{n} \sqrt{(\sum a_i \cdot \sum b_i)} \quad (1)$$

where considering the samples $\boldsymbol{a}$ and $\boldsymbol{b}$, $\boldsymbol{n}$ is the number of partitions, and $\boldsymbol{a}_i$, $\boldsymbol{b}_i$ are the number of members of samples $\boldsymbol{a}$ and $\boldsymbol{b}$ respectively in the $\boldsymbol{i}$'th partition. The Bhattacharya coefficient will range from 0 to 1 where 1 represents the completely similar image and 0 indicates that there is no similarity between two images. The concept of normalization will be used in Bhattacharya coefficient; normalization is a process that changes the range of pixel intensity values. The purpose is to bring the image with a different intensity values into a range that is more familiar and similar to the senses which in this case, the ranges is brought to values between 0 and 1 inclusive.

2.1.2 Shape Analysis

Another major image feature is the shape of an object. In shape-based image retrieval, the similarities of the shapes represented in images are measured. Generally, there are two categories of shape descriptors: boundary based and region based. In the boundary based shape descriptor, the focused is on the closed curve that surrounded the shape. There are various models describing this curve such as polygons, circular arcs, chain codes, etc. Region based shape descriptor such as Moment Invariants and Morphological Descriptor give emphasize to the entire shape region or the materials within the closed boundary. The dataset used in this research is blood cell images, where the aim is to determine the number of round objects. As such, one way to describe such images is to calculate the circularity ratio of the objects in an image. It represents how a shape is similar to a circle. The result of area and perimeter of an object inside each image will be used to form a simple metric indicating the roundness of an object using the following formula:

$$\text{Roundness Metric} = \frac{4\pi \times area}{perimeter^2} \quad (2)$$

This metric is equal to 1 only for a circle and it is less than 1 for any other shapes. The discrimination process can be controlled by setting an appropriate threshold. In this study, the threshold of 0.75 has been used since all the objects or bubbles in blood cell images are not completely round.

2.1.3 Bag of Words (BoW)

The process of BoW starts with detecting local interest point. Local interest point detectors have the task of extracting specific points and areas from images which are invariant to some geometric and photometric transformations. For the detection of local interest point, Difference of Gaussians (DoG) is used in this experiment. DoG detector proposed by Lowe [16] is invariant to translation, scale, rotation, and illumination changes and samples images at different locations and scales.

Next, distinctive feature that characterizes a set of keypoints for an image is extracted. Scale Invariant Feature Transform (SIFT) proposed by Lowe [17] is used to describe the grayscale image region around each keypoint in a scale and orientation invariant fashion. Each detected region is represented with the SIFT descriptor using the most common parameter configuration: 8 orientations and 4 × 4 blocks, resulting in a descriptor of 128 dimensions.

Next step in implementation of bag of visual words is the codebook construction where the 128-dimensional local image features have to be quantized into discrete visual words. This step uses k-means clustering method, and use cluster center as visual vocabulary term. Upon identification of cluster centers, each image is represented as histograms of these cluster centers by simply counting the frequency of the words appear in an image. To accomplish this task, each feature vector in an image is assigned to a cluster center using nearest neighbor with a Euclidean metric.

2.2 Retrieval Module

In this module, certain measurement techniques were used to determine the similarities between the feature extracted from query image and images in the database. Upon identifying the similarities, the respective images from the database will be then retrieved as similar images to the query image.

In color analysis, the color histogram of query image is compared with the color histogram of images in the database using Bhattacharya coefficient. The resulting value is a number ranging from 0 to 1 based on the normalization algorithm.

The threshold of 0.97 has been set; as such the top 10 images in the database where their similarity ratios are greater than the threshold value are selected as similar images to the query image.

Likewise, the same approach is used to retrieve the top 10 similar images to the query image based on shape descriptor of images. In this case, the ratio of circular objects existed in an image is compared with same ratio of all other images in the database. Then, those images having ratios close to the query image's ratio by ±3 are retrieved as similar images. In BoW based image retrieval approach, PLSA model is applied on extracted BoW to identify similar images to the query image as illustrated in Fig.3. These processes are explained in following:

Learning Phase:

1) Initially, PLSA model is trained on the set of images in the database with visual words (BoW) as an input to learn both $P(z_k|d_i)$ and $P(x_j|z_k)$.
2) **While** not converge **do**
   a) E-Step: Compute the posterior probabilities $P(z_k|d_i, x_j)$
   b) M-Step: Parameters $P(x_j|z_k)$ and $P(z_k|d_i)$ are updated from posterior probabilities computed in the E-Step.

   End While

Testing Phase:

1) The E-step and M-step are applied on the extracted BoW of the query image by keeping the probability of $P(x_j|z_k)$ learnt from the learning phase fixed.
2) Calculate the Euclidean distance between $P(z_k|d_i)$ and $(z_k|d_{query})$.
3) Those images with closest distance to $P(z_k|d_{query})$ will be retrieved as similar images.

The unobservable probability distribution $P(z_k|d_i)$ and $P(x_j|z_k)$ are learned from the data using the Expectation –Maximization (EM) algorithm. $P(z_k|d_i)$ denotes the probability of topic z_k given in document d_i. $P(x_j|z_k)$ denotes the probability of visual word x_j in topic z_k.

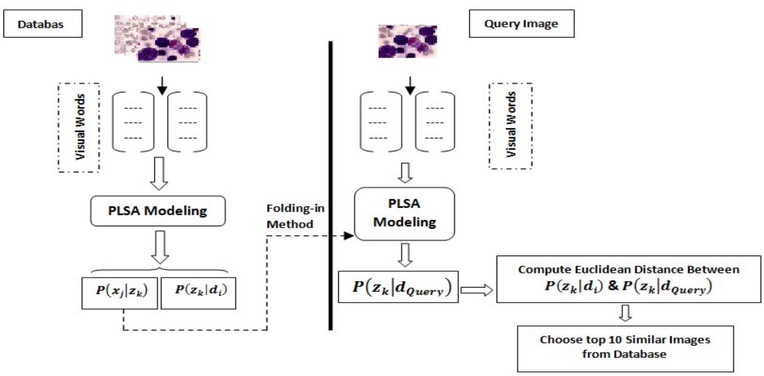

Fig. 1. Block Diagram of BoW-PLSA based Image Retrieval

Pseudo code of the retrieval process of 10 similar images to test images is demonstrated in following:

Start
1. Input $[BoW]_{600\times500}$ to PLSA model in order to compute $[XZ]_{500\times4}$ and $[ZD]_{4\times600}$
2. Insert the visual feature extracted from the unseen query image ($[BoW]_{1\times500}$) to PLSA model to compute matrix$[ZD]_{4\times1}$, by keeping $[XZ]_{500\times4}$ fixed from step 1.
3. The output is matrix$[ZD]_{4\times1}$.
4. *For i=1 to 600*

    Compute Euclidean distance between $[ZD]_{4\times i}$*and* $[ZD]_{4\times1}$.

    The top ten (10) vectors in matrix $[ZD]_{4\times i}$ *with closet distance to vector* $[ZD]_{4\times1}$ *will be selected. Each vector represents one image.*

    End For

End

3 Experimental Results and Discussion

Experiments were conducted to evaluate the retrieval performance obtained with respect to various image representation techniques. The database used in this research contains 600 blood cell images from four different classes; A1 Erythropoiesis, B1 Myeloid Cells Category, C1 Red Cell Disorder in the Neonate and Childhood , D1 Malarial Parasites Category. To evaluate the performance of the proposed retrieval system, we have randomly chosen 20 images from each class as query image. Precision is used to describe the accuracy of the proposed retrieval system.

Fig. 4 illustrates the average precision results obtained using the proposed retrieval approaches. The results showed that BoW-pLSA based retrieval approach outperformed the other two approaches.

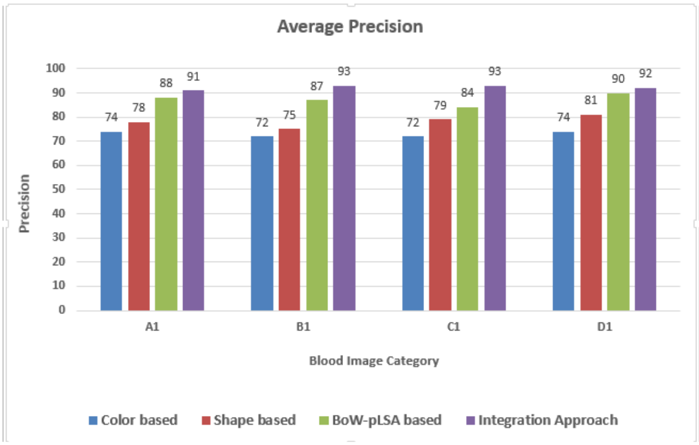

Fig. 2. Average Retrieval Precision

Analysis on the results shows that there are several irrelevant images retrieved using the color based approach which results in having lower retrieval precision. Further investigation on this results shows that most of these irrelevant retrieved images are having similar color histogram as the query image. This is due to the drawback of color histogram as it ignores the shape and texture of the images, i.e. two images from different category that happened to share color information can have similar color histograms. Fig. 5 illustrates this drawback; two images from Category B1 and one image from A1 category as well as their respective color histograms are presented.

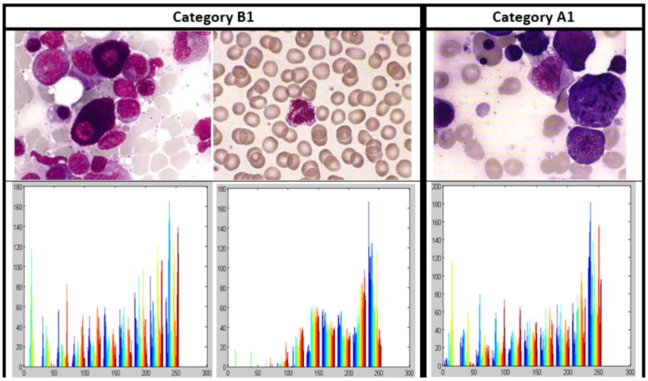

Fig. 3. Color Histogram of Sample Images from Two Categories

Comparing any retrieval systems is a difficult task especially in medical image retrieval domain. This is mostly due to the strong noises that exist in most of the medical images as well as intra-class variability and interclass similarities among them. In general, it is difficult to compare any two retrieval systems in the image retrieval domain. Due to the strong noise in most of the medical images as well as the existing similarities in the content of the images, it becomes imperative to use very precise descriptor. However, BoW with PLSA based image retrieval performs the best with average precision of 87.25% while shape and color analysis each has the precision of 78.25% and 73% respectively. The top 10 similar images to the query image retrieved from all the above three methods are analyzed separately. To increase the accuracy of retrieval results, only those images that retrieved in any of the above two methods were chosen as the final retrieval result. This leads the retrieval precision to be increased by 5%.

4 Conclusion

A wide availability of digital devices accelerates the growth of multimedia content production. Images are the most popular among the variety of multimedia contents. There is also an increase of digital information in medical domain such as blood cell images. As a result, there is an increased demand for a computerized system to manage these valuable resources. In this paper, various feature extraction techniques have been explored such as color and shape as well as bag of visual words. pLSA-BoW based image retrieval approach is also being exploited. A good performance has been obtained by combining the results obtained from the above feature extraction approaches.

References

1. Woo, C.S., Mirisaee, S.H.: Evaluation of a Content-Based Retrieval System for Blood Cell Images with Automated Methods. J. Med. Syst. 35, 571–578 (2011)
2. Hengen, H., Spoor, S., Pandit, M.: Analysis of Blood and Bone Marrow Smears using Digital Image Processing Techniques. In: SPIE Medical Imaging, San Diego, vol. 4684, pp. 624–635 (2002)
3. Cecilia, D.R., Andrew, D., Shahid, K.: Analysis of infected blood cell images using morphological operators. Image and Vision Computing 20, 133–146 (2002)
4. Pan, C., Yan, X., Zheng, C.: Recognition of blood and bone marrow cells using kernel-based image retrieval. International Journal of Computer Science and Network Security 6, 7 (2006)
5. Zare, M.R., Woo, C.S.: Integration of Color, Texture and Shape for Blood Cell Image Retrieval. Malaysian National Computer Confederation (2009),
6. `http://www.mncc.com.my`
7. Kesorn, K., Poslad, S.: An Enhanced Bag-of-Visual Word Vector Space Model to Represent Visual Content in Athletics Images. IEEE Trans. on Multimedia 14(1), 211–222 (2012)
8. Sui, L., Zhang, J., Zhuo, L., Yang, Y.C.: Research on pornographic images recognition method based on visual words in a compressed domain. IET Image Process 6(1), 87–93 (2012)
9. Zhou, W., Li, H., Lu, Y., Tian, Q.: Principal Visual Word Discovery for Automatic License Plate Detection. IEEE Trans. on Image Process. 21(9), 4269–4279 (2012)
10. Zare, M.R., Mueen, A., Woo, C.S.: Automatic Classification of Medical X-ray Images using Bag of Visual Words. IET Comp. Vision. 7(2), 105–114 (2013)
11. Dimitrovski, I., Kocev, D., Loskovska, S., Džeroski, S.: Hierarchical annotation of medical images. Pattern Recognition 44(10-11), 2436–2449 (2011)
12. Andre, B., Vercauteren, T., Buchner, A.M., Wallace, M.B., Ayache, N.: Learning Semantic and Visual Similarity for Endomicroscopy Video Retrieval. IEEE Trans. on Med. Imaging 31(6), 1276–1288 (2012)
13. Zare, M.R., Mueen, A., Woo, C.S.: Automatic Medical X-ray Image Classification using Annotation. J. Digit. Imag. (27), 77–89 (2014)
14. Yamaguchi, T., Maruyama, M.: Feature extraction for document image segmentation by PLSA model. In: DAS 2008 (2008)
15. Quelhas, P., Monay, F., Odobez, J.-M., Gatica-Perez, D., Tuytelaars, T., Gool, L.V.: Modeling Scenes with Local Descriptors and Latent Aspects. Paper presented at the Proceedings of the Tenth IEEE International Conference on Computer Vision (ICCV 2005), vol. 1 (2005)
16. Hofmann, T.: Unsupervised Learning by Probabilistic Latent Semantic Analysis. Mach. Learn. 42(1-2), 177–196 (2001)
17. Lowe, D.: Distinctive Image Features from Scale Invariant Key Points. International J. Comput.Vision 60(2), 91–110 (2004)
18. Lowe, D.G.: Object Recognition from Local Scale-Invariant Features. In: The Proceedings of the Seventh IEEE International Conference on Computer Vision (1999)
19. Blei, D.M., Ng, A.Y., Jordan, M.I.: Latent Dirichlet Allocation. J. Mach. Learn. Res. (3), 993–1022 (2003)