Empirical Comparison of Visual Descriptors for Content Based X-Ray Image Retrieval

Heelah A. Alraqibah[1], Mohamed Maher Ben Ismail[2], and Ouiem Bchir[2]

[1] King Abdulaziz City for Science and Technology, Riyadh, KSA
[2] College of Computer and Information Sciences, King Saud University, Riyadh, KSA

Abstract. Because of their visual characteristic which consists of black background versus white foreground, extracting relevant descriptors from medical X-ray images remains a challenging problem for medical imaging researchers. In this paper, we conduct an empirical comparison of several feature descriptors in order to evaluate their efficiency in content based X-ray image retrieval. We use a collection of X-ray images from Image-CLEF2009 data set in order to assess the performance of nine different visual descriptors with respect to different X-ray image categories.

Keywords: Visual descriptors, Content based image retrieval.

1   Introduction

X-ray images are the main source of information for several medical diagnosis processes. Thousands of digital medical images are produced everyday. Specifically, large-scale image databases containing various X-ray images are continuously growing, and computer systems such as Picture Archiving and Communication Systems (PACS) [1] have been proposed to overcome their storage and access challenges. Also, mining useful information from these databases have required efficient retrieval tools. Recently, Content Based Image Retrieval (CBIR) emerged as a novel component of medical diagnosis systems. CBIR consists of retrieving images based on the relevance of their visual content.Medical CBIR tools such as KMeD [2], COBRA [3], medGIFT [4], IRMA [5] have been proposed in order to enhance the performance of the medical diagnosis systems. One of the main phases of a typical content based image retrieval process is the visual features extraction. In fact, the visual content of an image is represented using one numerical vector which encodes its color, texture and shape properties. Recently, several studies have been conducted in order to improve the quality of extracted feature. For instance, Shim et al. [6] proposed an algorithm for X-ray image classification and retrieval using multi-class SVM with an ensemble feature vector by combining a color structure descriptor (CSD) based on the Harris corner detector and an Edge Histogram Descriptor of the image. In [7], the authors proposed a new approach for X-ray image clustering. They extracted the image feature vector at global level, and local patches of texture and shape feature. They used Gray Level Co-occurrence Matrix descriptor[12] as texture

descriptor. On the other hand, the shape feature is extracted using Canny edge operator [8] to create the edge histogram. In addition, they extracted pixel level features by resizing the images into 10 by 10 windows. The obtained image features from the three different levels for each X-ray image are concatenated to form a unique vector. Finally, the clustering is performed using the obtained vectors within the corresponding feature space. One of the critical components of a CBIR system consists in finding the appropriate image representation to extract the visual descriptors of the image collection. In other words, the retrieval accuracy of the system depends mainly on the efficiency of the extracted features. Most of the available CBIR systems use basic descriptors such as color, texture and shape [1,6,7,9]. Despite the efforts to propose efficient visual descriptors, no common descriptors have been used in the literature. The objective of the current work is to run an empirical investigation of the different visual features in order to assess their performances when used for content based image retrieval. Also, this work investigates the relevancy of these descriptors with respect to each X-ray image category. Namely, we investigate the Color moments [10], the Color Histogram [10], the Color Structure Descriptors [11], the Scalable Color Descriptors [11], the 2D wavelet descriptor [10], the Gabor filter descriptor [10], the Grey Level Co-occurrence Matrix [12], the Edge Histogram Descriptors [11], and a 7- dimensional region based shape descriptor [10].

2   Content Based Image Retrieval System

CBIR systems rely on the image content in order to search and retrieve relevant images to the user query. Fig. 1 shows a typical CBIR system. As it can be seen it includes an offline stage where feature vectors representing X-ray images are extracted and stored to form the feature vectors database. Then, these feature vectors are categorized into homogeneous clusters based on their visual descriptors. The clustering process minimizes the search space and reduces distances calculation cost during the query phase. Also, CBIR includes an online stage where the user launches the retrieval process by submitting a query image. Then, the query visual descriptors are extracted and compared to the visual descriptors of the clusters representative images obtained by clustering algorithm. Finally, the results are displayed to the user as an ordered list of images based on their similarity to the query. Recently, Some CBIR approaches incorporated machine learning techniques which learn features relevance, summarize data, and reduce the problem size [2,13,14]. Finding an appropriate image representation is a critical stage for such retrieval process. The accuracy of the image retrieval depends mainly on the choice of the appropriate visual features to be used to represent the images. The more these descriptors are discriminative, the better the results are.

2.1   Feature Extraction

Capturing the information embedded in X-ray images aims at bridging the gap between the visual content and the numerical representation of images. This
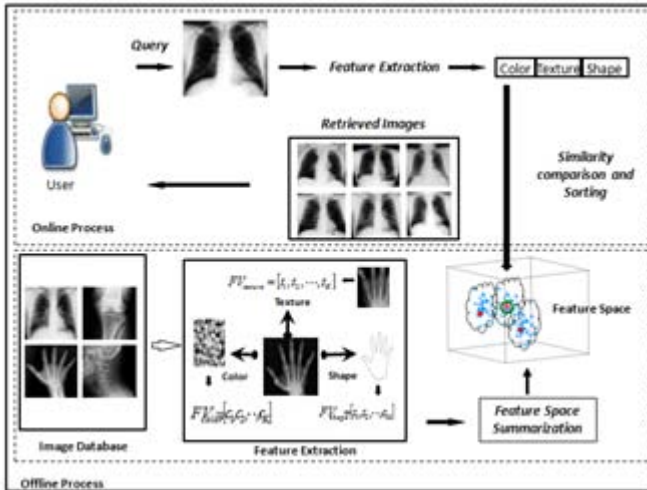
Fig. 1. Typical Content-Based Image Retrieval system

main step of any medical CBIR system is designed to encode the color, texture
and shapes present in the image. Color Feature is one of the basic descriptors
used for content based image retrieval. They are robust and stable compared
to other descriptors [11]. In the following we outline the set of color descriptors
which we investigate in this paper.

– Color moments Descriptor (CMD) [10] is a low-level feature which provides
  an efficient and effective representation of color distribution. The first order
  mean, the variance and the skewness are computed for the intensity values.
  The feature subset is represented using a 3-dimensional vector.
– Color histogram (CHD) [10] represents the number of pixels for each color
  bin. The color histogram calculation is robust to translation and rotation.
  The intensity histogram of an X-ray image is extracted using 256 bins.
– Color Structure Descriptor (CSD) [11] represents the local color structure of
  an image as histogram of local spatial structure of the color. It is extracted
  using an 8 by 8 structuring element which scans the image in a sliding window
  approach. Each time the structuring element is overlaid on the image, the CS
  histogram bins are incremented on the basis of the color present within the
  structuring element.
– Scalable Color Descriptor (SCD) [11] is a 256 bins histogram in the HSV color
  space. It uniformly quantizes and then encodes using Haar transform to reduce
  the bit number. This makes the application scalable.

The texture descriptors as a powerful discriminating low-level features contain
information of visual patterns in the image surfaces and their spatial location.
They are usually obtained using filter methods [15]. Below, we outline the texture

descriptors which we implement and assess their effectiveness on a medical CBIR system.

- Gabor Filter [10] is a Gaussian kernel function modulated by a sinusoidal envelop. A set of Gabor filters with different frequencies and orientations are used to extract useful features from the image. This consists of convolving the image with Gabor filters with different frequencies and orientations with respect to each frequency. Finally, the texture feature is represented by the mean and the standard deviation of the convolution result.
- Wavelet Texture Descriptor [10] is computed through recursive filtering and sub-sampling. The signal is decomposed into four different frequency sub-bands. Namely, these sub-bands are HH, HL, LH and LL, where L and H stand for low and high frequency, respectively. Usually, Daubechies' wavelet which is a linear combination of the wavelet function considering only the low-pass filter is used. Each image is decomposed into a number of levels, then, the feature vectors are building using the mean and standard deviation of each sub-band at each level.
- Grey Level Co-occurrence Matrix (GLCM) [12] contains information about occurrence of different combinations of gray levels values of pixel in an image considering the spatial relationship between pixels pairs. Each entry of GLCM corresponds to the number of occurrences of a pair of gray levels. Different co-occurrence matrix exists for each spatial relationship such as above, below, next to and etc. Different statistical features are extracted from these different matrices. Four co-occurrence matrices for different spatial relations (0, 45, 90 and 135) are computed for each image. Then contrast, energy, homogeneity and entropy features are extracted from the obtained co-occurrence matrices.
- Edge Histogram Descriptor (EHD) [11] captures spatial distribution of edges. The extraction of this descriptor consists in dividing the image into 16 non-overlapping blocks of equal size. Edge information is then calculated for each block in five edge categories: vertical, horizontal, diagonal, anti-diagonal and non-directional edge. The edge detector with the maximum edge value is then identified. If the edge value is above a given threshold, then the corresponding edge orientation is associated with the image-block. Also a global level of edge histogram is calculated. Another 13 semi global by grouping 13 different sub-block and generate edge distributions for five different edge types.

The Shape Descriptor is determined by seven geometrical properties of segmented region. Each input image is, first, segmented by simply binary conversion. Then, areas, eccentricity, extent, orientation, solidity, major axis length, and minor axis length of the obtained region are computed.

3   Experiments

We conducted our experiments on a collection of 1350 X-ray images containing 15 different categories. Namely, these categories are 1-Chest-supine, 2-Chest-sitting, 3-Chest side, 4-Skull, 5-Hand, 6-Spine, 7-Pelvis, 8-Spine side, 9-Skull
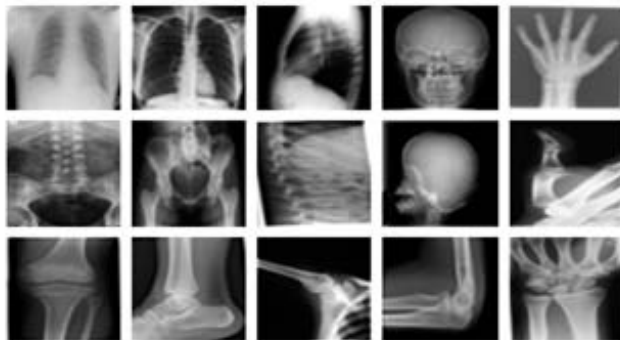
Fig. 2. Sample X-ray images from ImageCLEF2009

side, 10-Finger, 11-Knee, 12-Ankle, 13-Shoulder, 14-Elbow and 15-Wrist. This collection is a subset of the ImageCLEF2009 data set [16]. The images are classified according to IRMA code [7] which includes the class of the imaging modality, the examined body part, the image orientation with respect to the body, and the biological system. This information represents the ground for our experiments. Sample X-ray images from these categories are shown in Fig. 2.

The aim of our experiment is to study the effectiveness of different visual descriptors with respect to the different X-ray image categories. Therefore, we extract all the features outlined in section 2 from all images at the global level. Also, we extract these same features at the local level by dividing the original X-ray image into nine non-overlapping blocks. In order to assess the relevance of each feature with respect to each category, we use two approaches. The first one consists of computing the pair-wise Euclidean distances between pairs of images using the considered feature, and displaying their corresponding heat maps. The heat map gives an insight on the discrimination power of the feature. In each map, blue denotes a small distance and red represents maximal distance. In case the diagonal squares are almost deep blue while other squares turns red the feature considered discriminative. In Fig. 3 A, we show the heat maps of the extracted texture features at global level. As it can be seen, the performance of the different feature can vary significantly. For example, in Fig. 3.A.a, EHD descriptor performs very well in most of the categories except for two pairs of categories. Namely, the first pair is composed of class 4 and class 9 which correspond to skull and skull side images, respectively. The second pair consists of class1 "chest-supine" and class 2 "chest-sitting" categories. Further investigations showed that the corresponding categories share the same texture properties, and EHD cannot discriminate effectively between each class of them. The heat maps corresponding to the block based descriptors are shown in Fig. 3.B. It can be seen that the performances of all descriptors increased. The second experiment consists of using the considered features as input of the FCM clustering algorithm [18], and computes the accuracy of each obtained cluster. The accuracy is defined as the
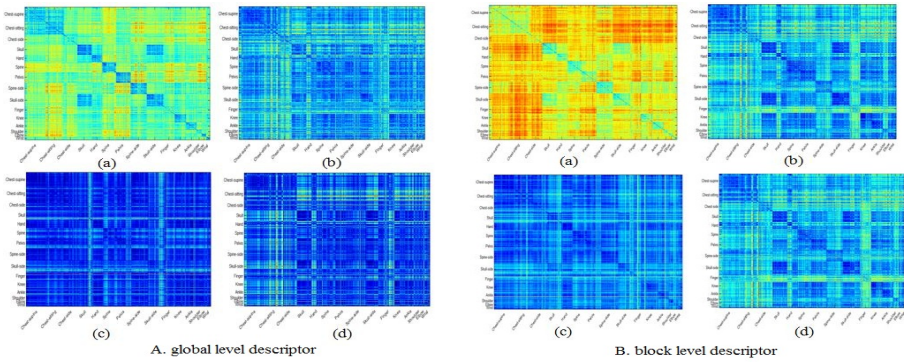
Fig. 3. The distances heat map for Texture descriptors:(a) EHD (b) Gabor filter (c) GLCM (d) Wavelet

ratio of the maximum number of images belonging to the same category by the total number of images in that category. This purity measure reflects the ability of the feature to discriminate between the different image categories, and discover the true structure of the data. In this experiment, we run FCM 15 times, and we consider the clustering partition corresponding to the highest obtained accuracy value. Fig. 4.a displays the obtained per-cluster purities obtained using FCM and the global descriptors. In fact, it shows that EHD descriptor outperforms the rest of the descriptors in all categories except the "finger" and "wrist" classes. For the "finger" class, CHD and CSD descriptors yield higher accuracy. On the other hand, for the "wrist" class, the SCD, shape and CHD descriptors yield the best performance measure. The shape descriptor gave better accuracy for "ankle" class compare to other features. The accuracy values obtained using FCM and block level descriptors are shown in Fig. 4.b. The performances of most descriptors increased. Table 1 shows the best accuracy performance with respect to each category. For instance, color histogram descriptor yields maximum accuracy for "finger" class at global level, while EHD has better accuracy at block level. On the other hand, "'wrist" category corresponds to the lowest clustering accuracy. This is due to the nature of the images in the data set; some images showing the wrist bone only while others contain the wrist bone, parts of the hand, and part of the forearm bones. The high performance of EHD is expected because according to the MPEG-7 standard, the EHD is very effective for content based image retrieval application. Moreover, the color descriptors have less performance because of X-ray images visual characteristics which consist of black background versus a white foreground.

4   Conclusion

In this paper we investigated different feature extraction methods in order to evaluate their effectiveness on different X-ray image categories. Also, we assessed the

Fig. 4. Accuracy of FCM clustering -15 clusters (a) global level descriptors (b) block level descriptors

Table 1. Per-cluster best accuracy and their corresponding descriptors

Category		Chest supine	Chest sitting	Chest side	Skull	Hand	Spine	Pelvis	Spine side
global	Feature	EHD	EHD	EHD	EHD	EHD	EHD	EHD	EHD
	Accuracy	0.55	0.64	0.83	0.86	0.86	0.89	0.83	0.65
block	Feature	GLCM	EHD	EHD	EHD	EHD	EHD	EHD	Gabor
	Accuracy	0.7	0.69	0.88	0.91	0.86	0.91	0.9	0.71
Category		Skull side	Finger	Knee	Ankle	Shoulder	Elbow	Wrist	
global	Feature	EHD	CHD	EHD	EHD	EHD	EHD	shape, SCD	
	Accuracy	0.71	0.79	0.9	0.82	0.76	0.82	0.4	
block	Feature	EHD	EHD	EHD	EHD	EHD	EHD	GLCM	
	Accuracy	0.77	0.62	0.9	0.92	0.92	0.82	0.47	

discrimination power of these visual descriptors using two different approaches. The first one relies on the analysis of the heat maps obtained by computing the pairwise distances between image categories in the corresponding feature space. The second approach consists in summarizing the X-ray image collection using FCM algorithm, and evaluating the obtained partition. The closer this partition is to the ground truth, the better the descriptor is. The experiment results showed that EHD outperforms the rest of the descriptors on most of the categories. Also, we conclude that the per-cluster accuracies of the clustering algorithm obtained using different descriptors can be improved if appropriate feature relevance weights are assigned with respect to each category. Thus, as future work, we intend to automatically learn these relevance weights. This can be achieved using SCAD [19] algorithm.

References

1. Akgül, C.B., et al.: Content-Based Image Retrieval in Radiology: Current Status and Future Directions. Journal of Digital Imaging 24, 208–222 (2011)
2. Chu, W.W., Cárdenas, A.F., Taira, R.K.: KMeD: A Knowledge-Based Multimedia Medical Distributed. Informatron Systems 20, 75–96 (1995)

3. El-Kwae, E.A., Xu, H., Kabuka, M.R.: Content-based retrieval in picture archiving and communication systems. Journal of Digital Imaging 13(2), 70–81 (2000)
4. Müller, H., Lovis, C., Geissbuhler, A.: The medgift project on medical image retrieval. University and Hospitals of Geneva, Service of Medical Informatics (2003)
5. Lehmann, T., et al.: Irma - a content-based approach to image retrieval in medical applications. In: Information Resources Management Association, pp. 911–912 (2006)
6. Shim, J., et al.: X-Ray Image Classification and Retrieval Using Ensemble Combination of Visual Descriptors. In: Third Pacific Rim Symposium, PSIVT, Tokyo, Japan (2009)
7. Ray, C., Sasmal, K.: A New Approach for Clustering of X-ray Images. International Journal of Computer Science Issues 7(4), 22–26 (2010)
8. Canny, J.: A computational approach to edge detection. IEEE Transaction Pattern Analysis and Machine Intelligence 8(6), 679–698 (1986)
9. Müller, H., Michous, N., Bandon, D., Geissbuhler, A.: A review of content- based image retrieval systems in medical applications clinical benefits and future directions. International Journal of Medical Informatics 73, 1–23 (2004)
10. Long, F., Zhang, H., Feng, D.D.: Fundamentals of Content-Based Image Retrieval. In: Multimedia Information Retrieval and Management. Springer
11. Manjunath, et al.: Color and Texture Descriptors. IEEE Transactions on Circuits and Systems for Video Technology 11(6), 703–715 (2001)
12. Ray, C., Sasmal, K.: A New Approach for Clustering of X-ray Images. IJCSI International Journal of Computer Science Issues 7(4), 22–26 (2010)
13. Yildizer, E., et al.: Integrating wavelets with clustering and indexing for effective content-based image retrieval. Knowledge-Based Systems 31, 55–66 (2012)
14. Rahman, M.M., et al.: A Framework for Medical Image Retrieval Using Machine Learning and Statistical Similarity Matching Techniques With Relevance Feedback. IEEE Transactions on Information Technology in Biomedicine 11(7), 58–70 (2007)
15. Arivazhagan, A., Benitta, R.: Texture classification using color local texture features. In: International Conference on Signal Processing Image Processing & Pattern Recognition ICSIPR (2013)
16. IRMA (Image Retrieval in Medical Applications), http://www.irma-project.org/
17. Lehmann, T., et al.: The IRMA code for unique classification of medical images. In: SPIE: Medical Imaging 2003: PACS and Integrated Medical Information Systems: Design and Evaluation (2003)
18. Maimon, O., Rokach, L.: Data Mining and Knowledge Discovery Handbook. Springer (2005)
19. Frigui, H., Nasraoui, O.: Simultaneous Clustering and Attribute Discrimination. In: The Ninth IEEE International Conference on Fuzzy Systems (2000)