X-Ray Image Classification and Retrieval Using Ensemble Combination of Visual Descriptors

JeongHee Shim, KiHee Park, ByoungChul Ko, and JaeYeal Nam

Dept of Computer Engineering, Keimyung University
1000 Shindangdong, Dalseo-Gu, Daegu, 704-701, Korea
{sjh0229,khp5500,niceko,jynam}@kmu.ac.kr

Abstract. In this paper, we propose a novel algorithm for the efficient classification and retrieval of medical images, especially X-ray images. Since medical images have bright foreground against dark background, we extract MPEG-7 visual descriptor from only salient parts of foreground. For color descriptor, Color Structure Descriptor (H-CSD) is extracted from salient points, which are detected by Harris corner detector. For texture descriptor, Edge Histogram Descriptor (EHD) is extracted from global and local parts of images. Then extracted feature vector is applied to multi-class Support Vector Machine (SVM) to give membership scores for each image. From the membership scores of H-CSD and EHD, two membership scores are combined as one ensemble feature and it is used for similarity matching of our retrieval system, MISS (Medical Information Searching System). The experimental results using CLEF-Med2007 images show that our system can indeed improve retrieval performance compared to other global property-based or other classification-based retrieval methods.

Keywords: H-CSD, EHD, SVM, ensemble vector, MISS.

1 Introduction

With the increase the digitalized medical images, various medical assistance systems, such as the Picture Archiving Communication System (PACS), have also been introduced that integrate information communication, computer networking, database management, and a user interface [1]. Therefore, the classification and retrieval of the medical images are important issue and the related studies are going on. The medical images have different meaning according to observer's viewpoints and consist of some interesting foreground regions and meaningless background. Therefore different classification and retrieval methods are required for medical images. Traditionally, medical images have been classified by experts and retrieved using just text. Yet, traditional classification and retrieval can produce irrecoverable mismatches according to the subjectivity and viewpoint of the experts. Furthermore, this kind of retrieval is costly and time consuming. Thus, to overcome these problems, various types of classification and Retrieval methods [2-4] have been proposed over the last few decades.

T. Wada, F. Huang, and S. Lin (Eds.): PSIVT 2009, LNCS 5414, pp. 738–747, 2009.

Mojsilovic et al. [2] proposed a method for semantic description, classification and retrieval of medical images. In this method, they used a semantic set of visual features, their relevance and organization for capturing the semantics of different image modalities. The Greenspan [3] represented images as some blobs with the Gaussian Mixture Model and estimated the matching scores between images using the KL (Kullback-Leibler). The Bhattacharya et al. [4] extracted the feature vectors using a color layer descriptor and a histogram descriptor of MPEG-7 standard descriptors. The extracted feature vectors were applied to SVM and FCM, which used to classify the medical images.

In this paper, we propose a novel algorithm for the classification and retrieval of the medical images. To classify medical images, we propose a Color Structure Descriptor (H-CSD) based on Harris corner detector for color feature and Edge Histogram Descriptor (EHD) for texture feature in the medical images. Then each extracted feature vector is applied to multiclass-SVM to give membership scores to each image. From membership scores of H-CSD and EHD, we combine two vectors into one ensemble vector and apply it to K-NI (K-Nearest Images) for image retrieval. Consequently, we can improve problems of the previous related works and provide the efficient method of the higher performance than the previous researches in classification and retrieval fields.

Rest of this paper is organized as follows: In Section 2, the algorithms for feature extraction using visual descriptors are described. The proposed classification and retrieval for medical image is introduced in Section 3. Section 4 evaluates the accuracy and applicability of the proposed classification method based on experiments, and some final conclusions and areas for future work are presented in Section 5.

2 Extraction of the Feature Values Using Visual Descriptors

To efficiently classify a lot of medical images into pre-defined categories, we first extract the feature vectors from images stored in database. In this paper, we use a CSD for color and an EHD for texture defined in MEPG-7 standard, respectively. Especially, CSD is modified to be extracted from only salient foreground regions using Harris corner detector and we named it H-CSD.

2.1 Color Structure Descriptor (CSD) Using Harris Corner Detector

Color is one of the most widely used visual features in image retrieval since it is relatively robust to viewing angle, translation and rotation of image. In this paper, we use the Color Structure Descriptor (CSD) to extract color vector because it aims at identifying localized color distributions using a small window. Furthermore it supports not only better the retrieval performance but also ease implementation than other color descriptors. The CSD is a descriptor represents an image by both the color histogram of the image and the local spatial structure of the color [5]. The elements of CSD are decided flexibly its size and the number of sub-sampling by the size of an image.

First, an image is quantized with 128 gray levels because X-ray image has only dark background and bright foreground. Then, image is divided into N x N sub-blocks. The size of sub-block is 8 x 8 pixels because MPEG-7 standard defines the scale of structuring element to be 8 x 8. In this paper, sub-block performs the same function as a structuring element. As we can see from Fig. 1, since our X-ray images contain useless background regions, we need to remove it before generating CS Histogram. Therefore we first detect Harris corner [6] from quantized image and only select sub-blocks containing one or more Harris points. The Harris corner detector is a popular point detector due to its strong invariance and stability against variation of viewpoint, illumination direction, scale and noise. From selected sub-blocks, 128-bin CS Histogram is extracted from an image represented in the 128-quantized gray color space. The CSD is a 1-D array of m bit-quantized values.

$$CSD = \overline{h}_s(m), \quad m \in \{1,...,M\} \tag{1}$$

where M is chosen from the set {256, 128, 64, 32} and where s is the scale of the associated structuring element (sub-block). At each position of sub-blocks, the CS Histogram is updated (accumulated) on the basis of the color present within the sub-block. For examples, if the eight gray levels and eight CS Histogram in 8 x 8 sub-block is created, the number of relevant bin increases 1 when the position of the CSD's sub-block on the image is corresponded to the pre-divided CS Histogram. In this way, the feature values of the image are the distribution of the number of the CSD's sub-blocks corresponding to each color histogram by the recorded color histogram. At final step, extracted 128-bin CSD is normalized to the range 0~1 for training of Support Vector Machine.

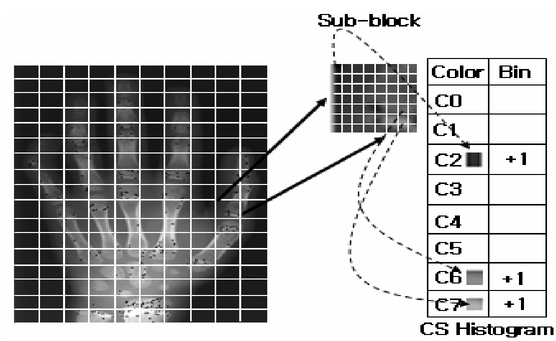

Fig. 1. Feature extraction process of H-CSD using Harris corner detector

2.2 Edge Histogram Descriptor (EHD)

EHD [7] is a descriptor that can represent the distribution of the regional edges of an image. Specially, dividing the image space into 4x4 non-overlapped sub-images and then each sub-image is further divided into non-overlapping square image blocks as shown in Fig. 2-(a). The local–edge distribution for each sub-image can be represented by a histogram. To generate the histogram, edges in the sub-images are categorized into five types; vertical, horizontal, 45 diagonal, 135 diagonal and non directional edges as shown in Fig. 2-(b). The size of the image block is decided by using equation (2) to divide input images into the same sized sub-images.

$$x = \sqrt{\frac{width \times height}{desired\ Num\ block}}, \qquad blocksize = \left\lfloor \frac{x}{2} \right\rfloor \times 2 \tag{2}$$

where the *desired Num block* is the whole number of the image blocks in the image. We decided default value as 1100 through experiments. Each of the image-blocks is then classified into one of the five edge categories mentioned above or as a non-edge block. If feature values are extracted by applying each filter, 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. Since there are 16 sub-images, a total of 5x16=80 histogram bins are generated.

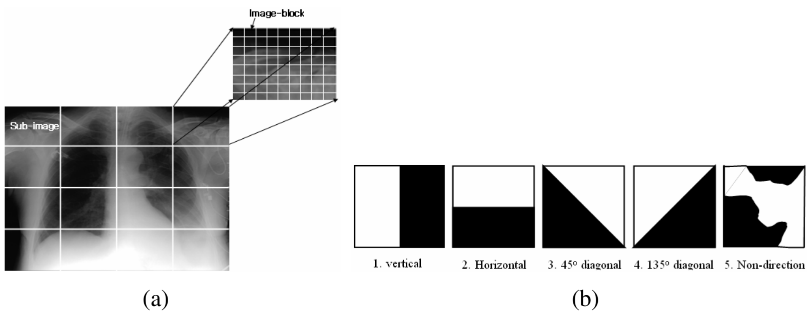

Fig. 2. (a) Definition of sub-image and image-blocks (b) five edge types for edge extraction

In classification and retrieval of medical images, especially, X-ray image, edge is the critical feature to improve the accuracy. Therefore we also extract five global edge histogram has 5 bins. Similarly, for semi-global edge histograms, we group 13 different subsets and generate edge distributions for five different edge types [7]. Finally, we use totally 150 edge histogram feature values by combining 80 regional edge histograms, 5 global edge histograms, and 65 (5 x 13) semi-local edge histograms.

3 Classification and Retrieval for Medical Images

After feature extraction, images are classified as one of predefined classes. To do this, we use multi-class Support Vector Machines (SVM) and two feature vector, H-CSE and EHD. By using training results, each image has membership scores on all 20 categories. These membership scores are estimated from H-CSD and EHD respectively and combined as one feature vector, *ensemble*. This ensemble feature vector is finally used for our content-based medical image retrieval system, MISS (Medical Information Searching System).

3.1 SVM Classification Using H-CSD and EHD Feature Vectors

An SVM can provide a good generalization performance for pattern classification problems without incorporating problem domain knowledge. Furthermore, an SVM does not require heuristic feature parameters for determining image classification.

Given training data $(\mathbf{x_1},...,\mathbf{x_N})$ that are vectors in space $\mathbf{x}_i \in \Re^d$ and their labels $(\mathbf{y_1},...,\mathbf{y_N})$ where $\mathbf{y}_i \in (+1,-1)^N$, the general form of the binary linear classification function is

$$\mathbf{g(x)} = \mathbf{w} \cdot \mathbf{x} + b \tag{3}$$

which corresponds to a separating hyperplane

$$\mathbf{w} \cdot \mathbf{x} + b = \mathbf{0} \tag{4}$$

where $\mathbf{x}$ is an input vector, $\mathbf{w}$ is a weight vector, and b is a bias. The main goal of SVM classifier is to find the parameter w and b for the optimal hyperplane that correctly separates the largest fraction of data points while maximizing the distance of either class from the hyperplane.

The SVM classification function is defined by [8]:

$$f(x) = sign(\sum_{i=1}^{l} v_i \cdot k(\mathbf{x}, \mathbf{x}_\mathrm{i}) + b) \tag{5}$$

where $k(\cdot\,,\cdot)$ is a kernel function, *vi* is weights for outputs of each kernel, b is a bias term and the sign of *f(x)* determine the class membership of x such as *+1* class and *-1* class. The decision function *f(x)* from the hyperplane determined by the support vectors can be used to measure how much an image belonging to the one category (+1) is different from the other categories (-1). Intuitively, the farther away a point is from the hyperplane, i.e. a larger positive $f(x)$, the more reliable the classification result.

For a linear SVM, the kernel function is just a simple dot product in the input space. However, in a non-linear SVM, the kernel function effectively projects the samples to a feature space of higher dimension **F** and constructs a hyperplane in **F** [8]. The SVM training algorithm then estimates a hyperplane that separates the data in **F** into two classes using the largest margin.

In this paper, we use multi-class SVM with RBF (Radial-Basis Function) Kernel instead of binary SVM because our x-ray images should be classified 20 classes according to regions of body. There are several commonly used methods, such as one-against-all, one-against-one, and directed acyclic graph [9]. Here we adapt the one-against-all method, which constructs n SVM classifiers where n is the number of classes. The *i-th* SVM is trained using all of the examples in the *i-th* class with positive labels (+1) and all others with negative labels (-1).

To perform the training, 1,754 images were randomly selected from 20 image categories as shown in Table 1. We used X-Ray images of IRMA (Image Retrieval in Medical Applications) which were used for Image CLEF med2007 [10].

In this paper, since we use two feature vectors respectively for performing training, *2n* SVM classifiers are generated.

Table 1. Training classes and images per one class for SVM

Category	Body Part	#of training data	Category	Body Part	#of training data
1	Breast	100	11	Finger	100
2	Pelvis	100	12	Wrist	100
3	Front head	100	13	Kneepan	100
4	Side head	100	14	Shoulder	100
5	Throat	100	15	Vertebrae	100
6	Knee	100	16	Front breast	100
7	Toe	48	17	Side breast	100
8	Front ankle	100	18	Fleshy	47
9	Side ankle	100	19	Elbow	21
10	Hand	100	20	Foot	38

3.2 Ensemble Feature Vector Combination and Similarity Matching

After SVM training, all database images having feature vectors of H-CSD and EHD are fed to the corresponding SVM classifiers and category membership scores are obtained at the output. In Figure 7, extracted feature vectors, F_c and F_E (F_c : feature vector of H-CSD, F_E : feature vector of EHD) of one image are fed to 2n SVM(2x20) classifiers respectively. Then SVM classifiers output 20 membership scores, $\vec{S}_c$ and $\vec{S}_E$, for each feature vector. Finally, the ensemble vector, $\vec{E} = [s_{c1}, s_{c2}, \ldots s_{c20}, s_{e1}, s_{e2}, \ldots, s_{e20}]$ is obtained by appending all category membership scores.

Originally, the test example *x* is fed into these *i-th* SVM classifiers and the one with the highest output score (*dj*) is selected as the final class.

$$d_j(x) = \max_{i=1,\ldots,n} d_i(x) \tag{6}$$

where *di* is the output score about *i-th* class for input *x* images. However, we combine output scores of 2n SVM classifiers as one ensemble vector. This ensemble vector is fed to final K-Nearest Neighbor cluster to predict the most likely top *k* categories for the given image. To retrieve most similar top k images from top k categories, the final distance is estimated by (7) and the top nearest images are displayed in ascending order of the final distance.

$$S(q,t) = \sum_{i=1}^{2n} | s_i^q - s_i^t | \tag{7}$$

where *q* and *t* denote query and target image.

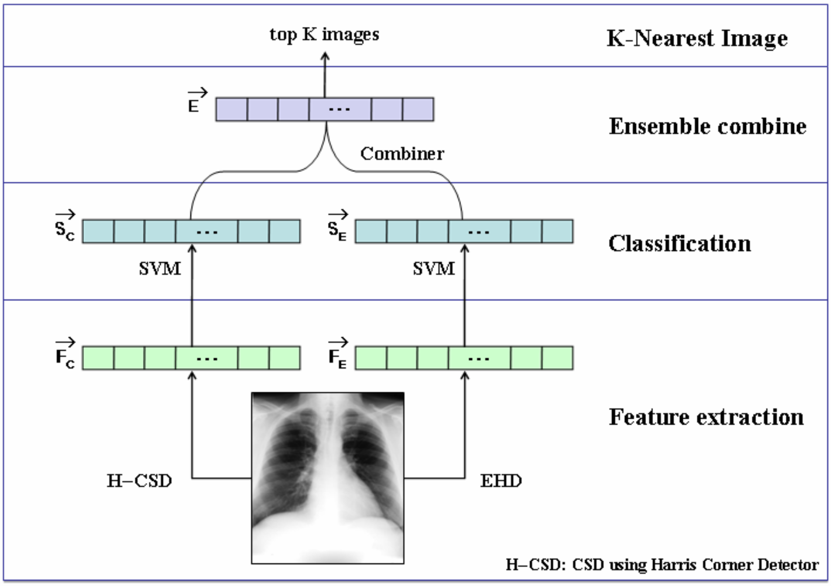

Fig. 3. Flow diagram of the classification and retrieval process

4 Experimental Results

The proposed system was developed using Visual C++ 6.0 language for off-line training and test system is developed based on ASP.NET 2.0 using C# language. For the test, we also use 1,000 images (20 categories) in IRMA (Image Retrieval in Medical Applications) [10]. Table 2 shows 20 categories for test and the number of test images for experiment. You can demonstrate our MISS system at our web-site, http://cvpr.kmu.ac.kr

Table 2. The titles of twenty categories and number of images for Test

Category	Body Part	#of test data	Category	Body Part	#of test data
1	Breast	60	11	Finger	50
2	Pelvis	60	12	Wrist	50
3	Front head	50	13	Kneepan	60
4	Side head	50	14	Shoulder	50
5	Throat	50	15	Vertebrae	60
6	Knee	50	16	Front breast	60
7	Toe	45	17	Side breast	50
8	Front ankle	50	18	Fleshy	40
9	Side ankle	50	19	Elbow	20
10	Hand	60	20	Foot	35

To complete a query, the user pushes the 'default' button and selects one retrieval method among eight methods. After that, the user clicks any image that he/she wants to retrieve and the top 20 nearest neighbors are returned. Figure 4 shows the retrieval interface of MISS.

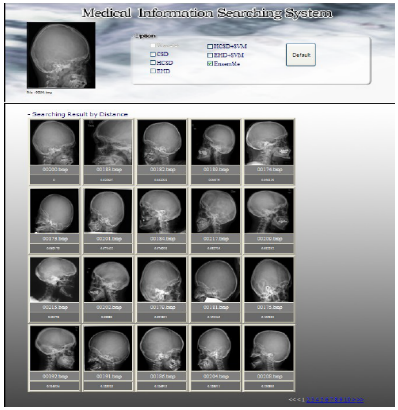

Fig. 4. Retrieval interface of MISS

To validate the effectiveness of our approach, we first compare the retrieval precision of our system with three methods which using only individual feature vector with SVM and similarity matching method.

The test is performed on 20 categories and 5 query images from each category. In all experiments, performance is measured using average retrieval precision. As shown in Table 3, the overall performance of our approach outperforms the other three methods as by percentages of 42.2%, 69.8%, 77.3%, and 96.5%. Especially, the retrieval performance of the proposed feature H-CSD showed a 27.6% improvement over the original CSD.

Table 3. The experimental results using descriptors independently

	Top = 5	**Top = 10**	**Top = 20**	**Average Precision**
CSD+SVM	0.52	0.426	0.32	0.422
H-CSD+SVM	0.744	0.693	0.658	0.698
EHD+SVM	0.81	0.77	0.7415	0.773
Ensemble	0.976	0.968	0.952	0.965

We also compared the retrieval performance with Bhattacharya et al. [4]'s algorithm. The Bhattacharya et al. [4] combined a color layer descriptor and EHD of MPEG-7 standard descriptors as one feature vector and applied it to SVM and FCM (Fuzzy C-mean Clustering). After that, the output scores of SVM and membership scores of FCM are linearly combined for classifing and retrieval the medical images. As we can see from Table 4, Bhattacharya's method shows the retrieval performance of an average 58.8%. In contrast, our proposed method showed a 37.7% improved retrieval performance. Figure 5 shows retrieval results of the MISS system.

Table 4. The experimental results using combination descripotrs

	Top = 5	Top = 10	Top = 20	Average Precision
Bhattacharya's method [SVM+FCM]	0.63	0.605	0.53	0.588
Ensemble vector	0.976	0.968	0.952	0.965

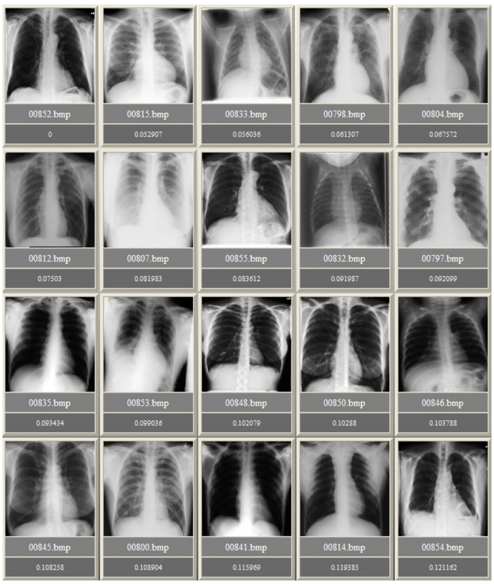

Fig. 5. Retrieval results using the proposed method about 'Front breast' category

5 Conclusion

In this paper, we proposed a novel algorithm for the efficient classification and retrieval of medical images, especially X-ray images. To classify medical images, we first extracted proposed Color Structure Descriptor (H-CSD) based on Harris corner detector for color feature. For texture descriptor, Edge Histogram Descriptor (EHD) was extracted from global and local parts of images. Then extracted feature vector was applied to multi-class Support Vector Machine (SVM) to give membership scores for each image. From the membership scores of H-CSD and EHD, ensemble one feature vector was generated and it was used for similarity matching of our retrieval system, MISS (Medical Information Searching System).The experimental results using CLEF-Med2007 images showed that our system could indeed improve retrieval performance compared to other global property-based or other classification-based retrieval methods.

In future works, improved algorithms for category classification and automatic annotation based on image classification are needed. Especially, we need to develop new feature to improve the classification performance on similar categories such as throat against vertebrae and finger against toe.

Acknowledgments. This work was supported by grant RTI04-01-01 from the Regional Technology Innovation Program of the Korean Ministry of Commerce, Industry, and Energy (MOCIE).

References

1. Qi, H., Snyder, W.E.: Content-based image retrieval in PACS. Journal of Digital Imaging 2, 81–83 (1999)
2. Mojsilovc, A., Gomes, J.: Semantic based categorization, browsing and retrieval in medical image databases. Int.Conf. on Image Processing 3, 145–148 (2002)
3. Greenspan, H.: Medical Image Categorization and Retrieval for PACS Using the GMM-KL Framework. IEEE Transactions on Information Technology in BioMedicine 11, 190–202 (2007)
4. Bhattacharya, P., Rahman, M.M.: Image Representation and Retrieval Using Support Vector Machine and Fuzzy C-means Clustering Based Semantical Spaces. In: International Conference on Pattern Recognition, vol. 2, pp. 1162–1168 (2006)
5. Manjunath, B.S., Salembier, P., Sikora, T.: Introduction to MPEG-7. John Willy & Sons, LTD. (2002)
6. Harris, C., Stephens, M.J.: A combined corner and edge detector. In: Alvey Vision Conference, pp. 147–152 (1998)
7. Won, C.S., Park, D.K.: Efficient Use of MPEG-7 Edge Histogram Descriptor. ETRI Journal 24, 23–30 (2002)
8. Vapnik, V.: The Nature of Statistical Learning Theory. Springer, Heidelberg (1999)
9. Chen, S.-C., Murphy, R.F.: A graphical model approach to automated classification of protein subcellular location patterns in multi-cell images. BMC Bioinformatics 7, 1–13 (2006)
10. Deselaers, T.: The CLEF 2005 Automatic Medical Image Annotation Task. International Journal of Computer Vision 74, 55–58 (2007)