IMPROVED REGION BASED MEDICAL IMAGE RETRIEVAL SYSTEM USING MULTIPLE TEXTURE FEATURES

Neelima.N[1], Sreenivasa Reddy.E[2]

[1] Assistant Professor, *Department of Electronics and Communication Engineering, School of Engineering and Technology, Jain University, Bangalore.*

[2] *Professor, Department of Computer Science Engineering, Acharya Nagarjuna University, Guntur, India.*

ABSTRACT

Abstract In present days, the importance of Image Retrieval process is increasing dramatically in Medical Applications, Digital forensics, and Surveillance systems. This paper proposes an Improved Region based image retrieval system using multiple texture features. The experimental results present a comparative analysis of using multiple features. In this work Local Binary Pattern (LBP), HU moments, Chebyshev & Fourier descriptors are used with Canberra, city block and Minkowski distances. The results also present the improved method based on accuracy. The accuracy is calculated by the ratio of number of images displayed to the number of similar images in the database. Experiments are evaluated on dataset consists of 500 images with 10 categories, each of 50 similar images.

Keyword: *shape features, city block distance, minkowski distance, fuzzy c-means cluster.*

1. INTRODUCTION

Present scenario, image retrieval is an important topic in different applications such as medical systems, video and still images repositories, digital album in the internet and forensics department [1]. Content-based Image Retrieval (CBIR) is an important tool to retrieve the image from Database. CBIR systems finds the shape and color features of database images and retrieve the images based on the distance of query and other images. The speed, precision and efficiency are the important parameters in CBIR. In CBIR system, the main task is to find the similar medical images from database on presenting a query image as input and then extract the features such as color, texture and shape [2]. The features of query image are compared to the features of the medical database images [3].
Two different types of approaches are used in CBIR systems. They are content based image retrieval [4] and region based image retrieval [5]. In recent work, extracting the objects from given image using different feature vectors with segmentation is considered. Therefore inaccurate segmentation will get the poor retrieval efficiency. So, segmentation is an important role in CBIR systems. The work [6] introduced the region based image retrieval system. SNL used to separate the image into regions and then those regions of query and database images are matched based on IRM heuristic [7]. This approach is expensive and highly complex. Feature descriptor is also important step in CBIR systems. The feature descriptor is used to reduce the dimension of image. In the past decades, CBIR based single feature descriptor used to retrieve the images. Single features are unsuitable for image retrievals. These are less accuracy than the multiple feature descriptors. In recent work, the combined feature descriptor are Harris [8], Hessian, scale invariant, affine invariant features are used in region based image retrieval.

The block diagram of the proposed method is shown in figure 1.First for the query image the segmented to divide them into clusters. then the features are extracted. In our method multiple features are combined into a vector and these features are compared with similarity measure with the features extracted from the image database. In this paper, we propose an Improved CBIR system with multiple features, which uses the fuzzy c-means clustering. The structure of the paper is as follows. Section II describes the fuzzy c-means clustering. The combined feature descriptor is explained in section III. Different distance calculation is mentioned in section IV. Results are showed in section V. concluded in section VI.

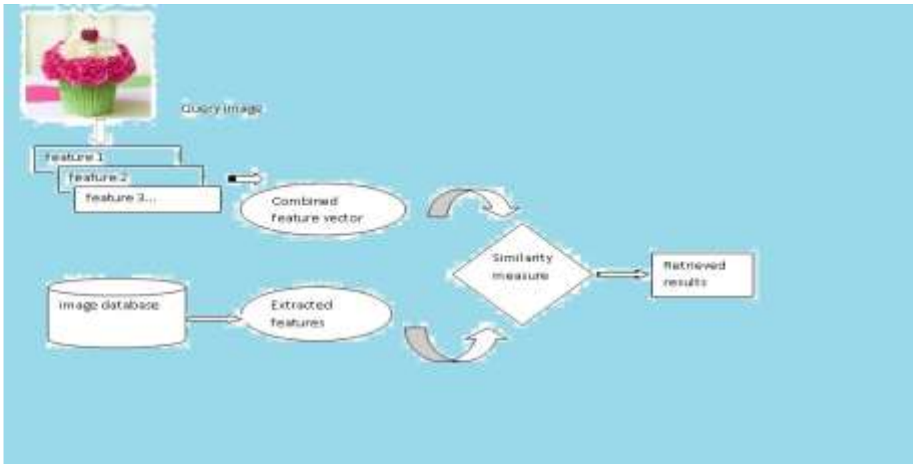

Fig.1 Block diagram of the proposed approach

2. FUZZY C-MEANS SEGMENTATION

The fuzzy c-mean clustering is an important key role in region based CBIR systems. The FCM is the process of grouping of image pixels into clusters by the appropriate similarity criterion such as distance and connectivity. It computes the cluster center using Gaussian weights. Four steps are followed in FCM algorithm. They are 1) fuzzy weights 2) standardize the weight over Q 3) eliminate empty cluster 4) update the cluster. These steps are described in section 2.1, 2.2, 2, 2.3 and 2.4.

2.1 Gaussian weights

Gaussian weight is also called fuzzy weights. In FCM, convert the image into vector and then choose initializing the random Gaussian weights. Computing the weights of vector in shown in equation 1.

$$w_{qk} = (1/(D_{qk})^2)^{1/(p-1)} \,/\, \Sigma_{(k=1,K)} (1/(D_{qk})^2)^{1/(p-1)}, \; p > 1 \quad (1)$$

where D is the image vector and p is the constant. Based on equation 1, assign a image vector value to a relative cluster depend upon maximum weight of the image vector over all clusters.

2.2 Gaussian weights over Q

In FCM cluster, computing the distance using Gaussian weights as shown in equation (2). It is closer to and grouping to one cluster.

$$w[q,k] = (w[q,k] - w_{min})/(w_{max} - w_{min}) \quad (2)$$

Where Wmin is the minimum Gaussian weight and Wmax is the maximum Gaussian weight of image vector.

2.3 Eliminating empty cluster

After computing distance, we add the Xie-Beni validity κ to avoid the empty cluster as shown in equation (3).

$$\kappa = D_{min2} / \{\Sigma_{(k=1,K)} \sigma_k^2\} \quad (3)$$

Where Dmin is the minimum value of image vector and σk2 can be given as

$$\sigma_k^{2} = \sum_{\{q:\ q \text{ is in cluster } k\}} w[q,k] \parallel \mathbf{x}(q) - \mathbf{c}(k) \parallel 2$$

All empty clusters are placed outside of image cluster loop. If the cluster is eliminated from the group then 0 is placed in cluster place.

2.4 Cluster Updating

To update the cluster centre and σk^2 of all vectors are shown in equation (4) & (5).

$$\nu = \{(1/K)\sum_{(k=1,K)} \sigma_k^{2}\}/D_{min}^{2} \quad (4)$$

$$\sigma_k^{2} = \sum_{(q=1,Q)} w_{qk} \parallel \mathbf{x}^{(q)} - \mathbf{c}^{(k)} \parallel^2 \quad (5)$$

3. COMBINED FEATURE DESCRIPTOR

In CBIR systems, feature descriptors are used to reduce the dimension of image. In this work, we combined multiple features as a single feature vector. They are image moments, Hu moment, local binary pattern, Chebyshev & Fourier descriptors. the combined feature vector has better retrieval accuracy than the single feature vector.

Image moments are used in segmenting the objects. Invariant moments (IM) are also called geometric moment invariants. Geometric moments, are the simplest of the moment functions with basis $\Psi_{pq} = x_p y_q$. It calculates the object centroid and average weights of object pixels. the moment function m_{pq} of order (p + q) is given as

$$m_{pq} = \sum_x \sum_y x^p y^q f(x,y) \quad \text{and} \quad \mu_{pq=\sum_x \sum_y (x-\overline{x})^p (y-\overline{y})^q} f(x,y) \quad (6)$$

Where f(x,y) is the pixel intensity of image object. The centroid of object in hu moment is given by $\bar{x}=m_{10}/m_{00}$ and $\bar{y}=m_{01}/m_{00}$

Local Binary Pattern is one type texture based feature vector. it is used to reduce the length of the feature vector and implement a simple rotation invariant descriptor. In images some binary patterns occur more commonly in texture images than others. A local binary pattern is called uniform if the binary pattern contains at most two 0-1 or 1-0 transitions. The local binary pattern (LBP) is given by

$$LBP_{P,R}(X_C, Y_c) = \Sigma S(g_p - g_c)\, 2^p \quad (6)$$

where g_p is the sample point of gray value with neighborhood of P sampling points and g_c is the gray level value of arbitrary pixel.

$g_p = I(x_p, y_p)$ $g_c = I(x, y)$. where $x_p = x + R\cos(2\pi p/P)$ and $y_p = y - R\sin(2\pi p/P)$.

Fourier descriptor is a power feature descriptor in image retrieval. Fourier descriptor is one of the low level shape feature. At low level frequency values contains information of shape features. at high level frequency values contain information of large shape details. At very high frequency values contain information of small shape details and not helpful in shape features. As a result, Fourier descriptor significantly reduces the size of image dimension.
Consider a set of coordinate of boundary shapes (x_k,y_k) where $0<k<N-1$.
The Fourier descriptor of boundary shape is given by

$$S(t)= X(t) + iY(t) \quad (7)$$

Chebyshev filter is a low level and shape feature in image retrieval. It highly reduces the image dimension. The Chebyshev filter of medical image size K × K is :

$P_o(x) = 1$ and $P_1(x) = (2x-K+1)/K$ where K is the radius point of Gaussian.

$t_p(x)= [(2m - 1)t_1(x)t_{m-1}(x) - (m - 1)(1-(m-1)/2)t_{m-2}(x)]/m$. where m range from 0 to infinite. The shape boundary coordinates X_k,Y_k is given by

$$X_k = rN/2(n-1)\cos(\theta) + K/2 \quad (8)$$

$$Y_k = rN/2(n-1)\sin(\theta) + K/2 \quad (9)$$

4. DISTANCE MEASUREMENTS

In this paper, we retrieve the similarity images based on different distances. They are Canberra, city block and Minkowski distances as explained in section 4.1, 4.2 and 4.3 respectively.

4.1 Canberra distance

The Canberra distance can be calculated between the feature vectors of query (Qi) and database images (Dbi) is given by

$$\text{Canberra}(Qi, Dbi) = \Sigma\, |Qi-Dbi| / [\,|Qi| + |Dbi|\,] \quad (10)$$

Where Qi= [Q1, Q2, Q3 …… Qi] Dbi= [Db1, Db2 … Dbi]

4.2 Minkowski distance

The Minkowski distance can be calculated between the feature vectors of query (Qi) and database images (Dbi) is given by

$$M(Qi, Dbi) = (\Sigma |Qi- Dbi|^{p})^{1/p} \quad (11)$$

Where Qi= [Q1, Q2, Q3 …… Qi] Dbi= [Db1, Db2 … Dbi]

4.3 Cityblock distance

The city block distance can be calculated between the feature vectors of query (Qi) and database images (Dbi) and is given by

$$\text{Canberra}(Qi, Dbi) = \Sigma\, |Qi-Dbi| \quad (12)$$

Where Qi= [Q1, Q2, Q3 …… Qi] Dbi= [Db1, Db2 … Dbi]
Finally the Accuracy is calculated by using the formula
Accuracy = number of images displayed/total no. of similar images in each set (50).

5. RESULTS

A dataset of about 500 Images are simulated, where realistic Computed Tomography (CT) and Magnetic Resonance images (MRI) are used. This work is simulated using MATLAB TOOL. The dataset consists of 500 images with 10 categories, each of 50 similar images. The database consists of similar images from 1 to 50 , 51 to 100, 101 to 150, 151 to 200, 201 to 250, 251 to 300 etc.

Fig.2 Input image

As an example the Fig.2 (image 101) is considered as the query input image, for which the similar images are to be retrieved. Fig.3 & Fig.4 displays the retrieval results of the input image with 5 features & 3 features in the combined feature vector using Minkowski distance. The accuracy is calculated based on the no. of images displayed. for image 101 the similar images in database are from 101-150. In Fig.3 the no. of images displayed in the range are 39 out of 50. In Fig.4 it is 50 out of 50. so the accuracy which is displayed in table 1 as 78% and 100%. The experiments are conducted by using various distances with different feature vectors. The results prove that the Minkowski distance has better efficiency compared to city block and Canberra distance.

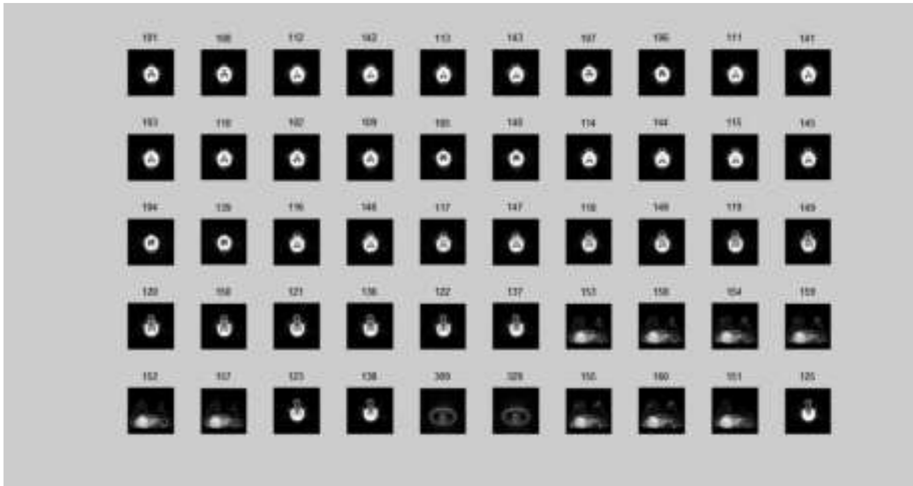

Fig.3 Retrieval results for the query image 101 using 5 features in feature vector

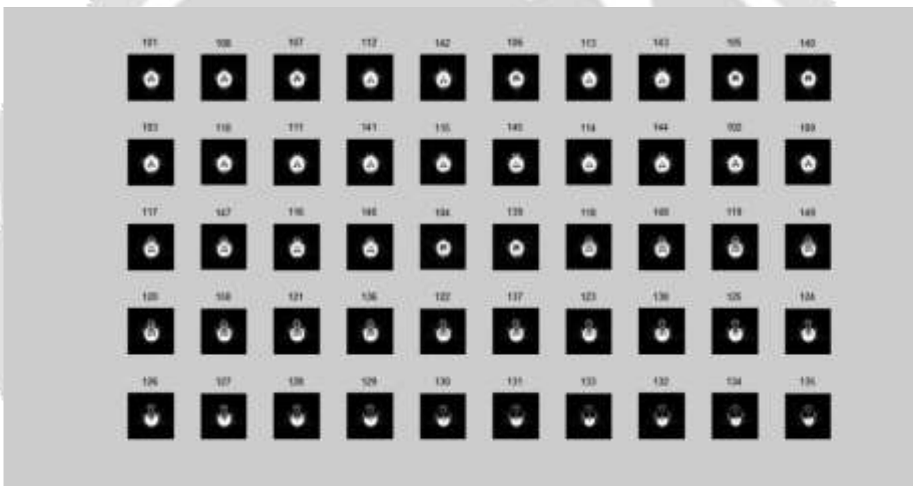

Fig.4 Retrieval results for the query image 101 using 3 features in feature vector

The accuracy of different distances with different feature vectors are shown in table 1& table 2 for different query images. The table gives the comparison of multiple features with different types of distances. the feature vector can be selected with two, three, four (or) five features. Fig. 5 is another example of displaying similar images with query image 346. The comparison table clearly shows that Minkowski distance with three features (HU, LBP & Chebyshev) in the vector provides better similar images from the database.

Fig.5 Retrieval results for the Image 346 using 5 features in feature vector

Fig.6 Retrieval results for the Image 184 using 3 features in feature vector

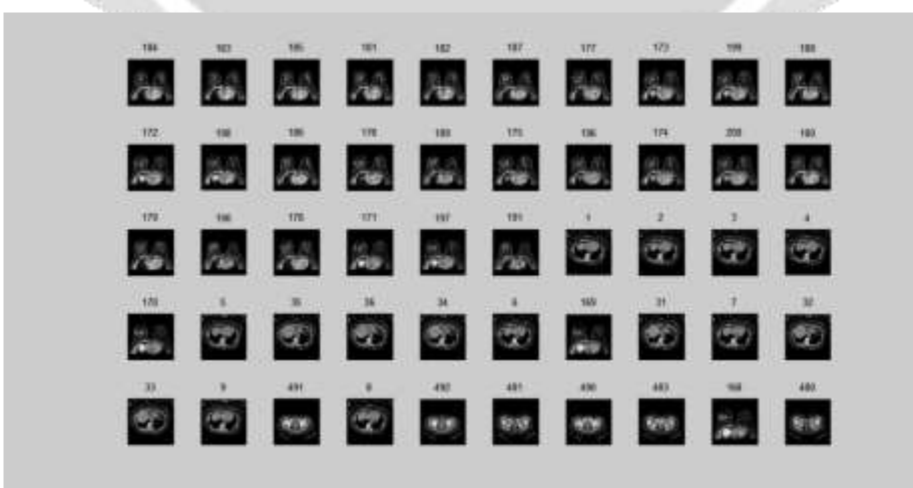

Fig.7 Retrieval results for the Image 184 using 5 features in feature vector

Let us consider another example with query image as image 184. For the query image the similar images in the database are from 151-200. Figure 6 & 7 are the results of three and five features in the combined feature vector. In figure 6 the number of similar images displayed are 50. In figure 7 the no of similar images displayed are 29 only. So the accuracies are 100 and 58 which are shown in table 3.

Feature vector / Distance	HU & LBP	HU,LBP& Chebyshev	HU, LBP, chebyshev & Fourier Descriptor	HU, LBP, Chebyshev , Fourier Descriptor, & Invariant moments
Minkowski	72%	**100%**	78%	78%
Canberra	66%	100%	68%	68%
City Block	84%	98%	90%	90%

Table - 1 Retrieval Accuaracy Comparision for the query Input Image 101

Feature vector / Distance	HU & LBP	HU,LBP& chebyshev	HU, LBP, Chebyshev & Fourier Descriptor	HU, LBP, chebyshev , Fourier Descriptor, & Invariant moments
Minkowski	54	**100**	58	58
Canberra	50	100	52	52
City Block	50	98	68	68

Table - 2 Retrieval Accuaracy Comparision for the query Input Image 184

Feature vector / Distance	HU & LBP	HU,LBP& Chebyshev	HU, LBP, Chebyshev & Fourier Descriptor	HU, LBP, Chebyshev , Fourier Descriptor, & Invariant moments
Minkowski	26	**92**	86	86
Canberra	22	90	88	88
City Block	48	92	88	88

Table - 3 Retrieval Accuaracy Comparision for the query Input Image 346

6. CONCLUSION

The paper presents the Improved Medical Image Retrieval System with multiple texture features. The experimental evaluation of the system is based on a 500 medical image database, which consists of ten categories of images each with fifty similar medial images. Experiments are conducted with different types of images which are presented in Table 1, 2 & 3. The experimental results prove that Minkowski distance, with three feature vectors (HU, LBP, Chebyshev) provides better retrieval Accuracy than Canberra and city block distances.

7. REFERENCES

[1]. Chih-Chin Lai, and Ying-Chuan Chen, "A User-Oriented Image Retrieval System Based on interactive Genetic Algorithm," IEEE transactions on instrumentation and measurement, vol. 60, no. 10, October 2011.

[2]. Rui, Y., Huang, T.S., Chang, S.: 'Image retrieval: current techniques, promising directions, and open issues', J. Vis. Commun. Image Represent, 1999, 10, (1), pp. 39–62.

[3]. Eakins, J.: 'Towards intelligent image retrieval', Pattern Recognition., 2002, 35, (1), pp. 3–14.

[4].Christina George Bab, Abraham Chandl: "content -based retinal image retrieval using dual-tree complex wavelet transform" 978-1-4673-4862-1/13/2013 IEEE.

[5]. Javeed Hussain, S., Satya Savithri, T., Sree Devi, P.V.: 'Segmentation of tissues in brain MRI Images using dynamic neuro-fuzzy technique', Int. J. Soft Comput. Eng., 2012, 1, (6), pp. 168–173.

[6]. V. Sridhar, M.A. Nascimento, X. Li, Region-based image retrieval using multiple-features, in: Proceedings of the 2002 Visual Information Systems Conference (VISUAL'02), 2002, pp. 61–75.

[7]. J. Li, J.Z. Wang, G. Wiederhold, IRM: integrated region matching for image retrieval, in: Proceedings of the Eighth ACM Multimedia Conference, 2000, pp. 147–156.

[8]. Harris, and M.J. Stephens, "A combined corner and edge detector," In Proceedings of Alvey Vision Conference, pp. 147-152, 1988.