2011 International Conference on Computer and Software Modeling
IPCSIT vol.14 (2011) © (2011) IACSIT Press, Singapore

Presented a New Method in Determining the Amount of Similarity Faces Using Symbols and HSV Colored Spectrum Histograms

Hossein Shirgahi[1,a] and Najmeh Danesh[2,b+]

[1] Jouybar Branch, Islamic Azad University, Jouybar, Iran

[2] Sari Branch, Islamic Azad University, Sari, Iran

Abstract. A lot of researches have been done about image processing these years. One of the most widely used fields is face recognition based on appearance features. In this research 250 samples of different people's faces has been taken in RGB mode as the input data and symbols and angles between applied symbols and considered parameters have been achieved and have been saved in the database. Then every image is changed to HSV and its histogram is made. Because of the importance of H spectrum, we considered 18 levels for its H's histogram and 3 levels for the histograms' of S and V. then we calculate the probability of histograms' levels of colored spectrum of gotten images and we register them. Finally, we determine the similarity of images by using symbols features and colored spectrums' histograms. As experimental results, the performance of this method is 12.5 percent more efficient than the similarity determination method based on symbol. Furthermore, this method is 19 percent more efficient than methods such as similarity determination method based on objective and spatial and similarity determination method which is only based on colored histogram.

Keywords: image processing, colored histogram, HSV colored mode, symbolic images.

1. Introduction

Image processing is one of the most widely used scopes in computer. This is a part of computerized images' scopes that these computerized processes are performed based on human's vision's system. For these kinds of usages, it is essential to know the quality of human's vision system. The main contexts of image processing's scopes are image retrieval, image improvement and image compression. Image retrieval is a most widely used contexts as analyses on images to reach some scientific, economical and security goals [6].

In most applications which are produced economically to do image retrieval for real environments, there are two important factors: running speed and accuracy of these applications. Also they are important in face recognition of image retrieval too. The most important parameter in produced software and hardware is accuracy, because image retrieval for face recognition is used in military and high secure environments. After accuracy, the speed of decision is important too. The previous works' accuracy is low because they considered only one or two features of an image.

R.R. Venkateswara and et al (2008) presented a new method for images' texture retrieval based on wavelets multi mode by Markov hidden tree. The features of images texture are extracted by using the effect of textures which are extended of sub groups' wavelets' coefficient. Experiments 'results show that this method reaches more punctual answers rather than previous methods which are based on wavelet in image texture retrieval [7].

+ [a]Hossein.Shirgahi@gmail.com, [b]Najmeh.Danesh@Gmail.com

In a research which has been done by P. Punitha and et al (2008), indexing on symbolic images is analyzed for retrieving similarity amount of a sample image to the images of a database. In this research, different shapes of each image are determined by the symbols. The indexes of each image are achieved by the distance and type of each image's symbols based on TSR algorithm. Finally a B+ tree is produced by all indexes of database's images and the amount of similarity is concluded based on new image's index values and searching in the B+ tree. On the basis of the conclusions, this method can calculate the image's similarities based on the considered symbols well [2].

Shu Ming Hsieh and et al (2008) perform the retrieval images based on spatial and objective similarities. They do it by object's conceptual information that is achieved of the images and spatial patterns of these objects. On the basis of performed researches, this method has more monotonous performance and dual speed rather than the previous methods, nevertheless the number of objects increases [3].

V.P. Subramanyam and et al (2007) presented a retrieval image system by using R tree. This method works based on visual descriptors of image like color, texture, description layout and image's shapes. Then they indexed the achieved information based on R tree. After that they presented a color fuzzy histogram to retrieve color and to gain the inaccuracies of descriptors of image's shapes. They performed the experiments on Corel images database. The performance of this system is 98 percent for different images of this database [5].

Up to now, those previous works used one or two features of an image for decision. By aware of this, we considered three parameters as symbol and color of an image to increase the accuracy of our method. Of course by increasing decision parameters, the speed of application's execution decreases to some extent. In this paper, to solve this problem we try to select methods which are in the same boat for extracting and analyzing symbol and color parameters of an image without doing extra operations.

We presented a new method in determining the degree of similarity using symbols and average color spectrum characteristics with color images and improved old methods in this scope [8].

2. Similarity's Amount Determination Based on Symbol

In this research, first we calculate each image's symbols by using edge detection algorithms and estimation methods and we saved them in a database as symbolic images. The symbolic images can be considered as structural description for physical images. Here when we say two images are the same, it means they are the same as color, shape, area and the sequence of three dimensional relations.

We use TSR model [2] to determine the amount of images similarity based on symbols. In this model, Quadruples (L_a, L_b, L_c, θ) shows the relations between elements of symbolic image. L_a , L_b and L_c are three symbols and θ is the angle between them. For every three symbols one Quadruple is made. As Fig. 1 shows, every six Quadruples can be made in exchange for every three elements A,B and C which are on one line. A situation that has one of the following conditions should be selected, because only one of the situations in exchange for every three symbols in the image should be considered:

$L_{i_1}, L_{i_2}, L_{i_3}$

labels are distinct and

$L_{i_1} > L_{i_2} > L_{i_3}$.

$L_{i_3} < L_{i_1},\ \ L_{i_1} = L_{i_2}$

$L_{i_1} > L_{i_2},\ \ L_{i_2} = L_{i_3}$

and

$Dist(comp(L_{i_1}), comp(L_{i_2}) \geq Dist(comp(L_{i_1}), comp(L_{i_3}))$.

$Dist(comp(L_{i_1}), comp(L_{i_2}) \geq M$

and

$L_{i_1} = L_{i_2} = L_{i_3}$

that M is calculated of Eq. 1.

$$M = Max(Dist(comp(L_{i_1}), comp(L_{i_3})), Dist(comp(L_{i_2}), comp(L_{i_3})) \quad (1)$$

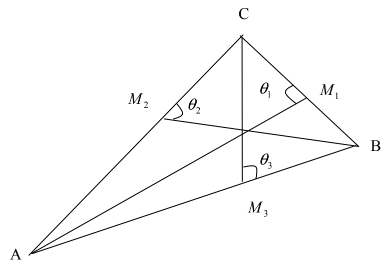

Fig. 1: triangular spatial relationship [2]

Dist is a function that expresses Euclidean distance between median points of each symbol. Comp is the symbol's center and θ is the smallest angle between element's median points. It is defined from Eq. 2.

$$\theta = \begin{cases} \theta_1 & if \quad \theta_1 < 90 \\ 180 - \theta_1 & othevwise \end{cases} \quad (2)$$

Of course if three symbols are on one line, θ is defined from Eq. 3.

$$\theta = \begin{cases} 90^\circ & if \quad L_{i_3} \quad comes \quad between \quad L_{i_1} \; and \quad L_{i_2} \\ 0^\circ & othevwise \end{cases} \quad (3)$$

Eq. 4 is used to show a unique number for Quadruple (L_a, L_b, L_c,θ) , then for all Quadruples of each image that are produced of this image's symbols, K_S values are calculated and are saved in database.

$$K_S = D_\theta(L_{i_1} - 1)m^2 + D_\theta(L_{i_2} - 1)m + D_\theta(L_{i_1} - 1) + (C_\theta - 1) \quad (4)$$

D_θ is the number of parts that is considered for the angle. In this research, the angle is in $[0^\circ..90^\circ]$ range and each degree is considered as a part, C_θ is the number of part which θ belongs to this part [2].

Finally to find the amount of similarity between an image of query (QI) and an image of database (DBI), first we calculate all K_SS of QI from Eq. 4, then we find the number of K_SS of DBI that is similar to it. So the amount of similarity based on symbol can be calculated from Eq. 5.

$$S_S(QI, DBI) = \frac{TNS_{K_s}(DBI)}{TN_{K_s}(QI)} \quad (5)$$

$TNS_{K_S}(DBI)$ is the number of K_SS of one DBI that is similar to K_SS of QI. $TN_{K_S}(QI)$ is the total number of K_SS of QI.

3. Similarity's Amount Determination Based

One of the features that are used to determine the amount of similarity is color. To determine this feature, we use color histograms. For an image, the color histograms are more general presentation of feature's vector of image's color. They are set of sub spaces and every histogram shows the probability of a special color of a pixel in the image. Histogram H is defined from Eq. 6 as vector.

$$H = \{H[0], H[1], ..., H[i], ..., H[N]\} \quad (6)$$

i is a quantized color in the color space, H[i] is the number of i color's pixels in every image and N is the number of sub spaces of color histogram[1].As researches done, HSV color space has a good harmonious with human's quality of perception of the color. Therefore in this step first we change the input RGB images to HSV, then we quantize this space to 162 distinct parts. To quantize it, we use 18 different values for H and 3 different values for S and V [4]. So, we make histograms for all color spectrum based on the quantization. A sample of it is shown in Fig. 2. Then we calculate the probability of each of H, S and V levels based on Eq.

7, 8 and 9 respectively. In these Equations, P_{HHi} is the probability of i's level of H spectrum, P_{Hsi} is the probability of i's level of S spectrum, P_{Hvi} is the probability of i's level of V spectrum, N_{HHi} is the number of H spectrum pixels that are in i' level, N_{Hsi} is the number of S spectrum pixels that are in i' level, N_{Hvi} is the number of V spectrum pixels that are in i' level and N_T is the total number of image's pixels.

$$P_{HH_i} = N_{HH_i} / N_T \quad 1 \le i \le 18 \tag{7}$$

$$P_{HS_i} = N_{HS_i} / N_T \quad 1 \le i \le 3 \tag{8}$$

$$P_{HV_i} = N_{HV_i} / N_T \quad 1 \le i \le 3 \tag{9}$$

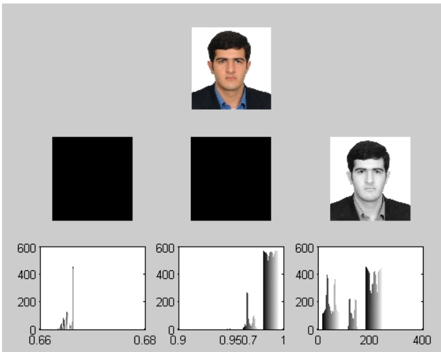

Fig. 2: H, S and V colored spectrums with histograms of a sample image.

Then we compare the probability of each level of colored spectrums point to point and the levels which their probability difference is less than 0.01 will be considered as similar levels. Based on these preamble, the amount of similarity between two images is calculated based on colored spectrum levels of H,S and V from Eq.s 10, 11 and 12. N_{SHH} ،N_{SHS} and N_{SHV} are similar levels in H,S and V spectrums respectively and N_{THH} ،N_{THS} and N_{THV} are total number of levels of V,S and H colored spectrums respectively. Finally the amount of similarity of two images based on color is calculated from Eq. 13. In this research, α_1 ،α_2 and α_3 coefficients determine the effects of S_{HH} ،S_{HS} and S_{HV} parameters respectively. Summation of these three coefficients should be one. Based on our experiments the best coefficients of them are 0.4, 0.3 and 0.3 respectively.

$$S_{HH} = N_{SHH} / N_{THH} \tag{10}$$

$$S_{HS} = N_{SHS} / N_{THS} \tag{11}$$

$$S_{HV} = N_{SHV} / N_{THV} \tag{12}$$

$$\begin{cases} S_H = \alpha_1 S_{HH} + \alpha_2 S_{HS} + \alpha_3 S_{HV} \\ \alpha_1 + \alpha_2 + \alpha_3 = 1 \end{cases} \tag{13}$$

4. Similarity's Amount Determination Based on Mixture of Symbol and Colored Histograms

After calculating the amount of similarity based on symbol (S_S) and the amount of similarity of colored spectrums' histograms (S_H), now we calculate the amount of total similarity (S_T) based on these two parameters from Eq. 14. β_1 and β_2 coefficients are the effects of S_S and S_H parameters respectively, the summation of β_1 and β_2 should be one. Based on our experiments, the best values for these coefficients are 0.6 and 0.4 respectively. It shows that the effect of image's texture is more than colored Histogram's.

$$\begin{cases} S_T = \beta_1 S_S + \beta_2 S_H \\ \beta_1 + \beta_2 = 1 \end{cases} \tag{14}$$

5. Experiments and Tests

We implemented this research by MATLAB 7 and SQL SERVER 2000 database software. First we save 250 images of different people in the database. The information of the image which has been saved included images' physical address, the symbols of each image with symbol's number and image's number, gotten K_S s for each image and the probability of each colored spectrums H,S and V levels.

Then for comparing a sample image to one or some other images we suppose that if the sample image is new and it is not in the database, we calculate its K_S s parameters and the probability of colored spectrums H,S and V levels. But if it is not a new one, we retrieve its desired parameters from the database on symbol to respective image of the database from Eq. 5. In the next step, we calculate the amount of similarity based on colored histograms of these two images from Eq. 13 based on the probabilities of colored spectrums H,S and V levels of the sample image and desired image of the database. Finally we calculate the total similarity from Eq. 14. There are some sample images in Fig. 3 and the amount of similarity of sample image 1 to sample images 2, 3 and 4 is expressed in Table 1.

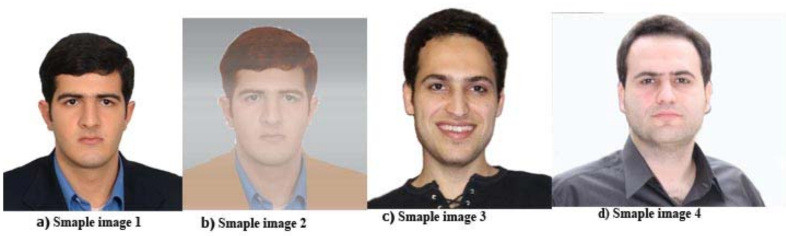

Fig. 3: Sample images

Table 1. The amount of similarity of sample image 1 with sample images 2, 3 and 4

Sample Image 1 compared with	similarity based on symbol	similarity based on colored Histogram	Ultimate similarity
Sample Image 2	98%	41.5%	75.4%
Sample Image 3	42.8%	51.1%	46.12%
Sample Image 4	51%	36%	44.9%

We compared our new method with method which calculates the amount of similarity only based on symbol [2], the method which calculates it based on spatial and objective similarities [3] and the method which calculates it only based on colored histogram. These comparisons are shown in Fig. 4. From this Figure, it is clear that our new method performs the same as the method which is only based on symbol for the images that color does not have such an effect. But for images that color has an important effect on the similarity, our new method's efficiency is 10 to 15 percent more. This method performs 15 to 23 percent better than the methods which are based on spatial and objective similarities and based on only colored histograms.

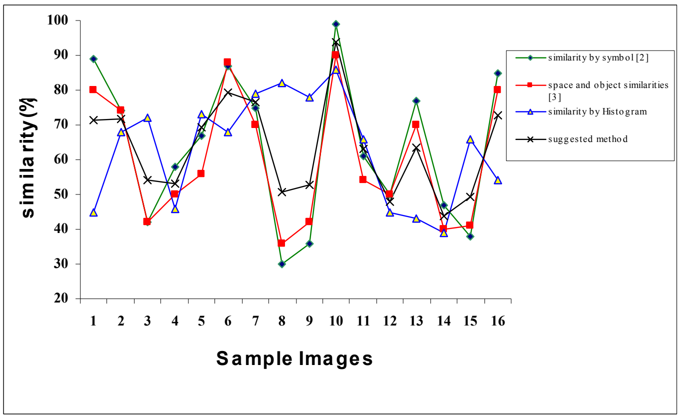

Fig. 4: Comparison between the new method and previous methods for determining the similarity's amount of images.

The time complexity of our new method is O(Log N), it is the same as the method which is only based on symbol [2], Because when we implemented this method, we used B+ tree like [3] for searching based on kss . Of course the access time and search time can be optimal by multi dimensional B^+ tree and R tree structures.

6. Conclusions

After performing tests and experiments, we reach empirical conclusions as bellow:

This proposed method is an efficient method and it can determine images' amount of similarity compared with the amount of real similarity with more than 95 percent accuracy.

This proposed method's efficiency is 12.5 percent more than the method which is only based on symbol to determine the amount of similarity.

This proposed method's efficiency is 19 percent more than the methods which are only based on colored histograms and the methods based on spatial and objective similarity.

As future works, in spite of using assurance ranges for matching images, we can use histograms of colored spectrums' fuzzy variables to determine the similarity amounts based on fuzzy rules. Furthermore, we can use symbol features and histograms of colored spectrums in addition to images' texture features to determine the amount of similarity. Another method to expand this, is image blocking to avoid overlapping of images' total average's side effects in histograms of colored spectrums and applying blocking decisions on all blocks based on our new method.

7. References

[1] R. Gonzalez and R. E. Woods. *Digital Image Processing*. 3rd edition, Addison Wesley publishing company, New York, USA, August 2007.

[2] P. Punithaa and D. S. Gurub: Symbolic image indexing and retrieval by spatial similarity:An approach based on B-tree. *Pattern Recognition*. 41, 2008, p. 2068 – 2085.

[3] S. M. Hsieh and C. C. Hsu: Retrieval of images by spatial and object similarities, I*nformation Processing and Management*. 44, 2008, p. 1214–1233.

[4] J. R. Smith and S. F. Chang: VisualSeek: A fully automated content-based image query system, *Proc. Int. Conf. on Image Proceeding*. 1996.

[5] V. P. Subramanyam and S. K. Sett: Image retrieval system using R-tree self-organizing map, *Data & Knowledge Engineering*. 61, 2007, p. 524–539.

[6] S. E. UMBAUGH: Computer Vision & Image Processing. Indianapolis, Books craft, 1998.

[7] R. R. Venkateswara and V.P. Subramanyam Rallabandi: Rotation-invariant texture retrieval using wavelet-based hidden Markov trees. *Signal Processing*.88, 2008, p. 2593– 2598.

[8] H. Shirgahi and N. Danesh. A new method in determining the degree of similarity using symbols and average color spectrum characteristics. *Indian Journal of Science and Technology*. 3(10), Oct 2010, p. 1094-1097.