10 years - ANNIVERSARY EDITION *TRAKIA JOURNAL OF SCIENCES*
Trakia Journal of Sciences, Vol. 10, No 4, pp 5-12, 2012

Available online at:
http://www.uni-sz.bg

ISSN 1313-7069 (print)
ISSN 1313-3551 (online)

Original Contribution

DETERMINE SIMILARITY OF FACES USING SYMBOLS AND AVERAGE COLOR SPECTRUM CHARACTERISTICS WITH FUZZY SYSTEM

H. Shirgahi[1*], N. Danesh[2]

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

ABSTRACT
One of the most widely used areas in the field of image processing is face recognition based on appearance characteristics, in this study to determine its similarity to the face, first we capture 200 samples of images of different people as RGB color mode and we convert this images to HSV color mode and we quantized HSV color mode to 32 levels, then we obtains the symbols of each image with the average color of each symbol. We calculate desired parameters based on angles places being created symbols and average color spectrum and stored in our database. And then based on the angle between the symbols used in the files, the desired level and parameters of color symbol Fuzzy set of files and ultimately the level of similarity are obtained based on defined fuzzy rules. Based on the observed experimental results, this method compared to determine similarity based on the symbol is 12 percent more efficient.

Key words: Image Processing, average Color Spectrum, HSV Color Mode, symbolic Images, fuzzy System

INTRODUCTION
Digital image processing is a method to apply computerized algorithms in order to process the digital images in routine usages. Image processing is a part of computerized images scopes in that this computerized processes are performed based on human vision system. For these kinds of usages, it is essential to know the mechanism of human vision system. The main contexts of image processing 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, commercial and security goals (1).

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 too. The most important parameter in producing software and hardware is accuracy, because image retrieval for face recognition is used in military and high security environments. After accuracy, the speed of decision making is important too. In the previous studies 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 using 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 gives more accurate answers rather than the previous methods which are based on wavelet in image texture retrieval (2).

In a research done by P. Punitha and et al (2008), indexing on symbolic images was analyzed for retrieving the amount of similarity between a sample image and 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 the image of each symbol based on TSR algorithm.

***Correspondence to**: *Hossein Shirgah, Faculty Department of Computer, Islamic Azad University, Jouybar Branch, Computer Software Engineering. Post code: 477761-861310; Telephone: +98-09113261904; Fax: +98-124-3363190.*

Finally a B^+ tree is produced by all indexes of database 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 (3).

Shu Ming Hsieh and et al (2008) performed image retrieval based on spatial and objective similarities. They retrieve images based on the conceptual information obtained in the object images and spatial patterns. On the basis of performed researches, this method has more monotonous performance and dual speed rather than the previous methods, despite an increase in the number of objects (4).

V.P. Subramanyam and et al (2007) presented a retrieval image system using R tree. This method works based on visual descriptors of image like color, texture, description layout of the 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 gain the inaccuracies of the descriptors of image shapes. They performed the experiments on Corel images database. The performance of this system is 98 percent for different images of this database (5).

Being aware of this, we considered two parameters of symbol and colored spectrums average of an image to increase the accuracy of this method. Of course the speed of running on application decreases to some extent by increasing decision parameters. In this paper, to solve this problem we tried to select methods which are in the same boat to extract and analyze the parameters of symbol and color of an image without doing extra operations. As everyone knows, if the condition of taking pictures changes a little, the pictures of that object may change a lot. To avoid these kinds of problems, we used fuzzy system to define the amount of similarity and fuzzy deciding instead of logical deciding in this paper.

DETERMINING SIMILARITY AMOUNT BASED ON SYMBOL

In this research, first we calculate the symbols of each image by using edge detection algorithms and estimation methods and saved them in a database as symbolic images. The symbolic images can be considered as structural description of physical images, **figure 1** shows a sample of symbolic images. Here when we say two images are the same, it means they are the same not only in color, shape and area but also in the sequence of three dimensional relations.

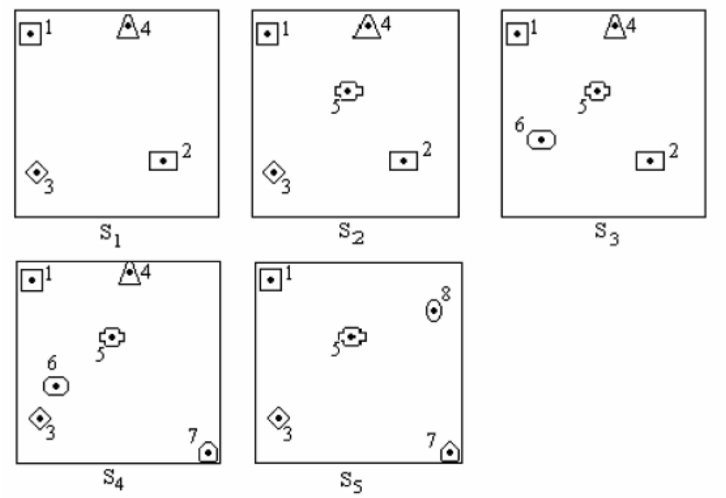

Figure 1. A sample of symbolic images (3)

TSR model (3) is used to determine the amount of image similarity based on symbols. In this model, Quadruples (L_a, L_b, L_c, θ) shows the relations between the elements of symbolic image. L_a , L_b and L_c are the three symbols and θ is the angle between them. For every three symbols one quadruple is made. As **figure 2** shows, every six Quadruples can be made in exchange for every three elements A, B and C which are on one line. A situation should be selected that has one of the following conditions, because only one of the situations in exchange for every three symbols in the image should be considered:

1- Labels $L_{i_1}, L_{i_2}, L_{i_3}$ are distinct and $L_{i_1} > L_{i_2} > L_{i_3}$.

2- $L_{i_3} < L_{i_1}$, $L_{i_1} = L_{i_2}$

3- $L_{i_1} > L_{i_2}$, $L_{i_2} = L_{i_3}$ And $Dist(comp(L_{i_1}), comp(L_{i_2}) \geq Dist(comp(L_{i_1}), comp(L_{i_3}))$.

4- $Dist(comp(L_{i_1}), comp(L_{i_2}) \geq M$ and $L_{i_1} = L_{i_2} = L_{i_3}$ that M is calculated using g formula 1.

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

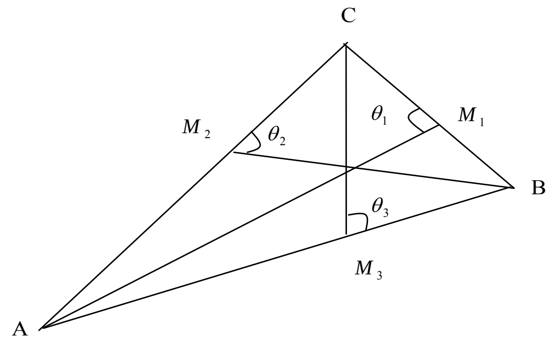

Figure 2. Triangular spatial relationship (3)

Dist is a function that expresses Euclidean distance between median points of each symbol. Comp shows the center of the symbol and θ is the smallest angle between the median points of element. It is defined according to formula 2.

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

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

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

Formula 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, KS values are calculated and are saved in database.

$$K_S = D_\theta(L_{i_1} - 1)m^2 + D_\theta(L_{i_2} - 1)m + D_\theta(L_{i_3} - 1) + (C_\theta - 1) \quad (4)$$

D_θ is the number of parts considered for the angle. In this research, the angle is in $[0^\circ..90^\circ]$ range and each degree is considered as one part, C_θ is the number of the part to which θ belongs (3).

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

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

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

DETERMINING THE AMOUNT OF SIMILARITY BASED ON COLORED AVERAGE

One of the features used to determine the amount of similarity is color. To determine this feature, we used colored average. For an image, the colored average is the simplest presentation of the vector of the feature of image color(6). According to the research, HSV color space has a good harmony with the way humans perceive color. Therefore in this step first we change the input RGB images to HSV, then we quantize this space to 1024 distinct parts. To quantize it, we use 16 different values for H and 8 different values for S and V.

You can see a sample image with H, S and V image spectrums in figure 3. The obtained value based on H, S and V image spectrums average is one of the comparative parameters in the similarity of an image. This parameter is not a suitable criterion to solve some problems like colored spectrums values overlap lonely. the amount of similarity of an image of query (QI) with an image in the database (DBI) is calculated based on the average of each H, S and V colored spectrums according to 6, 7 and 8 formulas. Here L is a function which returns the amount of quantization of each spectrum.

$$S_{HM}(QI_H, DBI_H) = \frac{\min(L(DBI_H), L(QI_H))}{\max(L(DBI_H), L(QI_H))} \qquad (6)$$

$$S_{SM}(QI_S, DBI_S) = \frac{\min(L(DBI_S), L(QI_S))}{\max(L(DBI_S), L(QI_S))} \qquad (7)$$

$$S_{VM}(QI_V, DBI_V) = \frac{\min(L(DBI_V), L(QI_V))}{\max(L(DBI_V), L(QI_V))} \qquad (8)$$

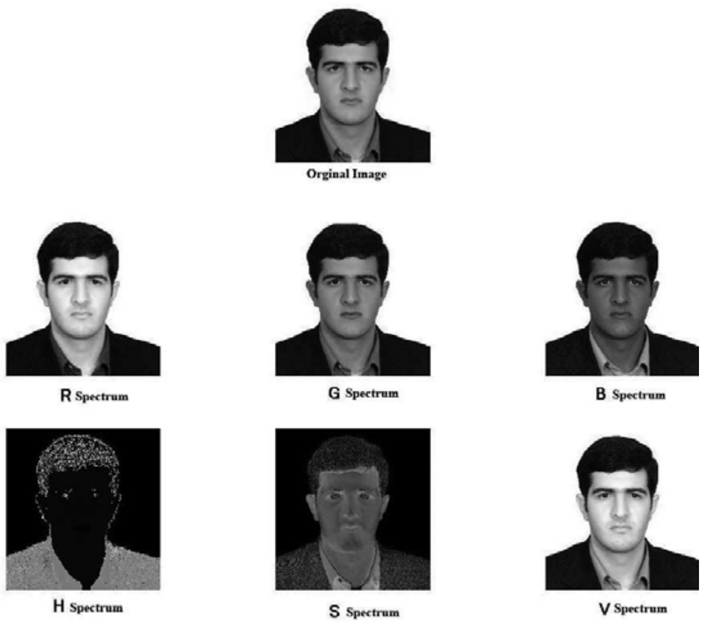

Figure 3. RGB and HSV colored spectrums of a sample image

DETERMINING SIMILARITY AMOUNT BASED ON THE FEATURES OF SYMBOL AND COLORED SPECTRUMS AVERAGE BY FUZZY SYSTEM

After calculating the amount of similarity based on symbol (SS) and the amount of similarity of colored spectrums average (SH), now we improve the method of the amount of similarity (ST) based on symbol.

We use fuzzy system to determine the amount of similarity based on the incorporation of symbol characteristics and colored spectrums average. Therefore, we consider the fuzzy variables of the amount of similarity based on symbol and the amount of similarity based on colored spectrums H,S and V.

Fuzzy sets of each variable are shown in **figure 4, 5, 6 and 7.** Note that fuzzy sets of these variables are based on parabola function.

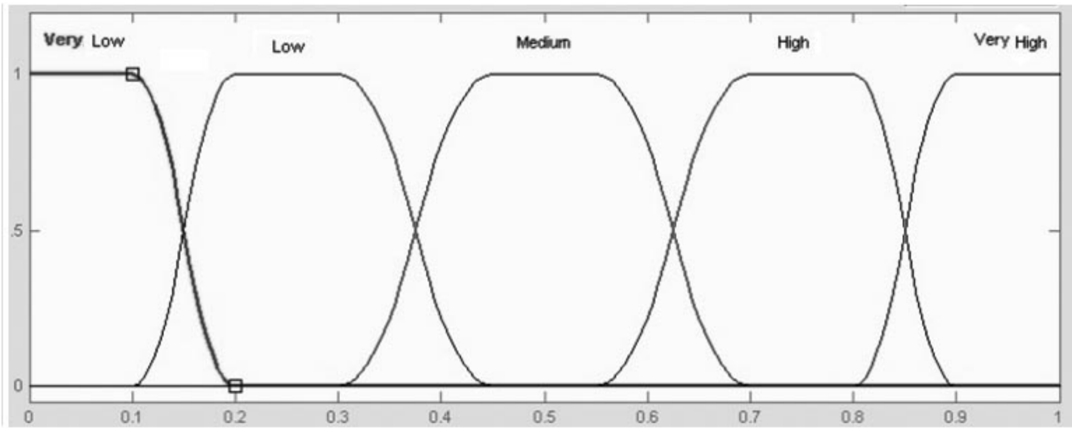

Figure 4. Fuzzy sets of fuzzy variables of the amount of similarity based on symbol

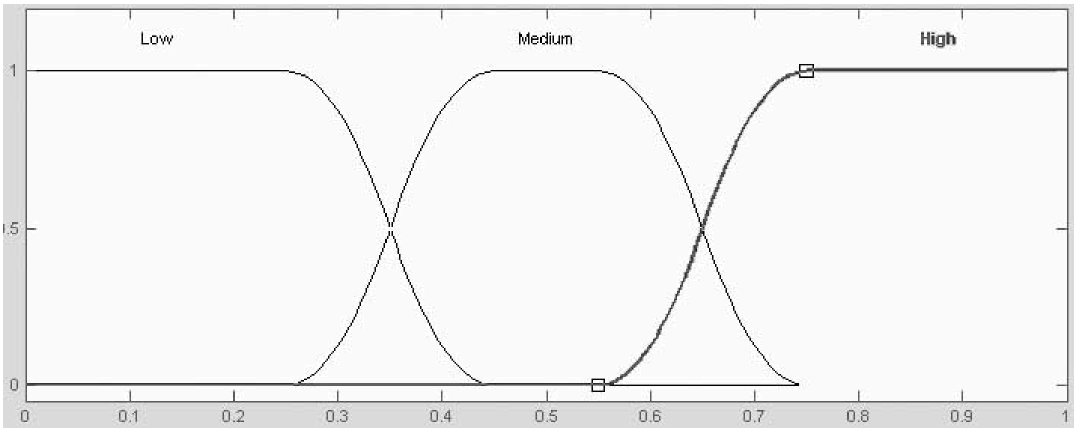

Figure 5. Fuzzy sets of fuzzy variables of the amount of similarity based on H spectrum

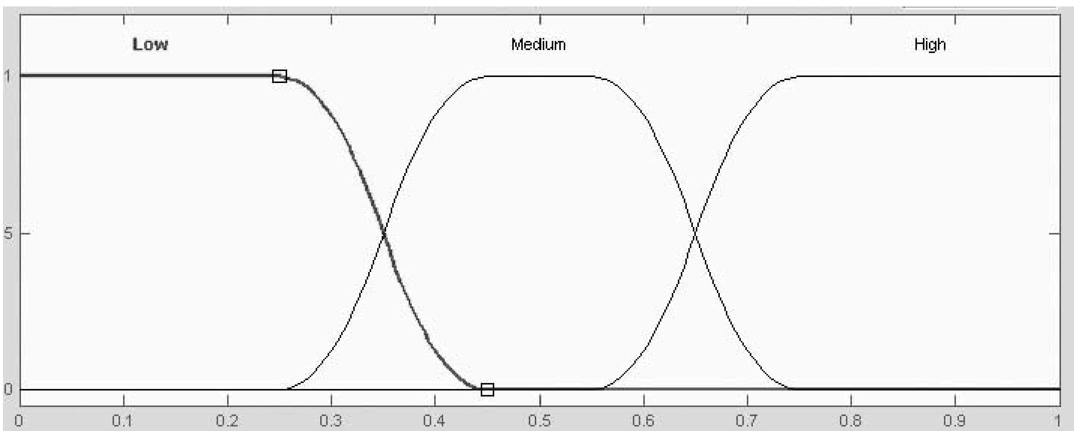

Figure 6. Fuzzy sets of fuzzy variables of the amount of similarity based on S spectrum

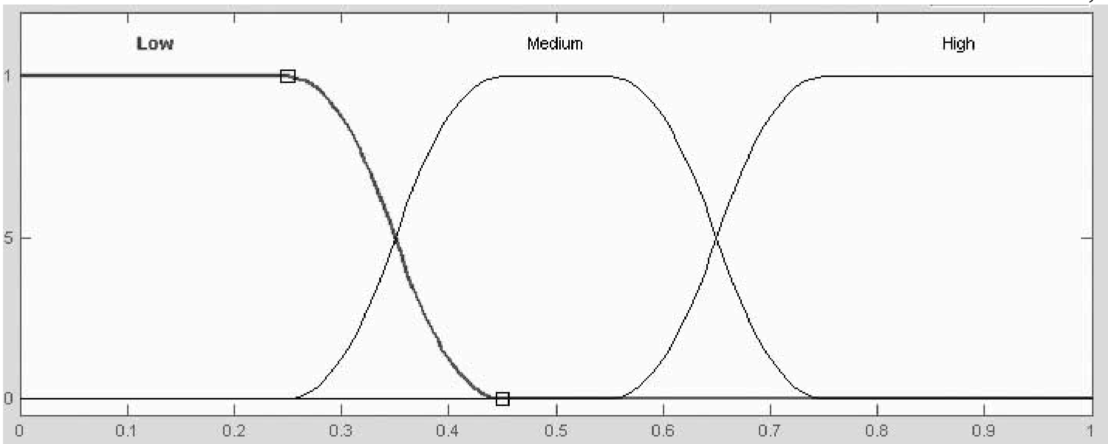

Figure 7. Fuzzy sets of fuzzy variables of the amount of similarity based on V spectrum

After determining the fuzzy sets of input variables, we determine an output fuzzy variable called the amount of similarity. Its fuzzy sets are shown in **figure 8**. Note that the fuzzy sets of output variable is based on parabola function too.

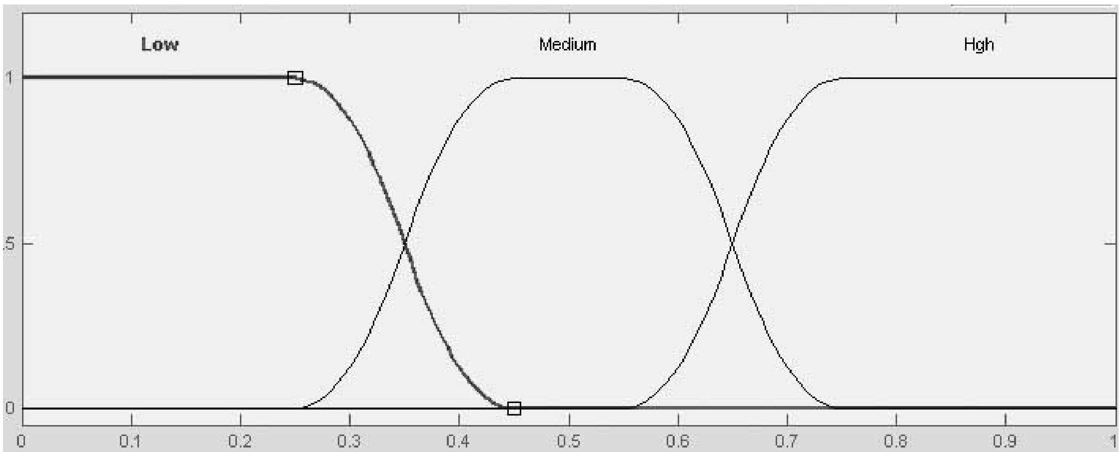

Figure 8. The fuzzy sets of output fuzzy variable

After implementing input and output fuzzy variables and their fuzzy sets, we determine fuzzy rules based on the amount of similarity according to incorporation of symbol characteristics and colored spectrums average and also the amount of each input parameter's dependency and their effects on output variable. In other words, we determine the relation between input and output variables by fuzzy rules. You can see some fuzzy rules in **figure 9.**

We consider more probability for the fuzzy variable of the amount of similarity based on symbol in fuzzy rules, because its effect is more. To express the rules in fuzzy sets we use max product inference on fuzzy sets in fuzzy rules (7).

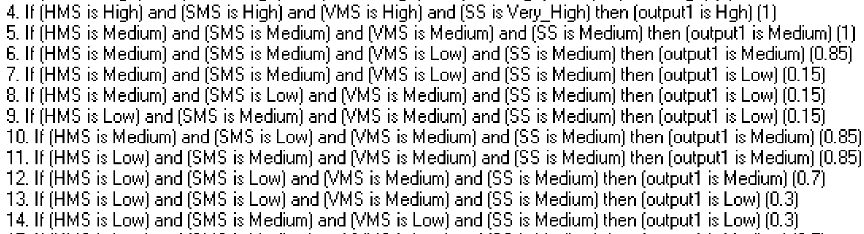
1. If (HMS is Low) and (SMS is Low) and (VMS is Low) and (SS is Very_Low) then (output1 is Low) (1)
2. If (HMS is Low) and (SMS is Low) and (VMS is Low) and (SS is Low) then (output1 is Low) (1)
3. If (HMS is High) and (SMS is High) and (VMS is High) and (SS is High) then (output1 is Hgh) (1)
4. If (HMS is High) and (SMS is High) and (VMS is High) and (SS is Very_High) then (output1 is Hgh) (1)
5. If (HMS is Medium) and (SMS is Medium) and (VMS is Medium) and (SS is Medium) then (output1 is Medium) (1)
6. If (HMS is Medium) and (SMS is Medium) and (VMS is Low) and (SS is Medium) then (output1 is Medium) (0.85)
7. If (HMS is Medium) and (SMS is Medium) and (VMS is Low) and (SS is Medium) then (output1 is Low) (0.15)
8. If (HMS is Medium) and (SMS is Low) and (VMS is Medium) and (SS is Medium) then (output1 is Low) (0.15)
9. If (HMS is Low) and (SMS is Medium) and (VMS is Medium) and (SS is Medium) then (output1 is Low) (0.15)
10. If (HMS is Medium) and (SMS is Low) and (VMS is Medium) and (SS is Medium) then (output1 is Medium) (0.85)
11. If (HMS is Low) and (SMS is Medium) and (VMS is Medium) and (SS is Medium) then (output1 is Medium) (0.85)
12. If (HMS is Low) and (SMS is Low) and (VMS is Medium) and (SS is Medium) then (output1 is Medium) (0.7)
13. If (HMS is Low) and (SMS is Low) and (VMS is Medium) and (SS is Medium) then (output1 is Low) (0.3)
14. If (HMS is Low) and (SMS is Medium) and (VMS is Low) and (SS is Medium) then (output1 is Low) (0.3)

Figure 9. Some defined fuzzy rules

After expressing fuzzy rules, we should perform defuzzification operation to achieve the final result. In this paper we use fuzzy centroid defuzzification algorithm according to formula 9 (7).

$$y_i = \frac{\sum_{j=1}^{P} y_j * m_{B'}(y_j)}{\sum_{j=1}^{P} m_{B'}(y_j)} \quad (9)$$

EXPERIMENTS AND TESTS

We implemented this research by the software of MATLAB 7. First we save 200 images of different people in the database. Here we use SQL SERVER 2000 as a database system. The information saved about the images are the physical address of images, the symbols of each image along with the symbol number and image number, gotten K_s obtained for each image. Then in order to compare a sample image to one or some other images we suppose that if the sample image is new and it is not available in the database, we calculate the parameters K_s based on formula 6. But if it is available in the database, we retrieve its desired parameters from the database. Finally we calculate the ultimate similarity based on expressed fuzzy system. There are some sample images in **figure 10** and the amount of similarity of sample image 1 to sample images 2, 3 and 4 is shown in **Table 1.**

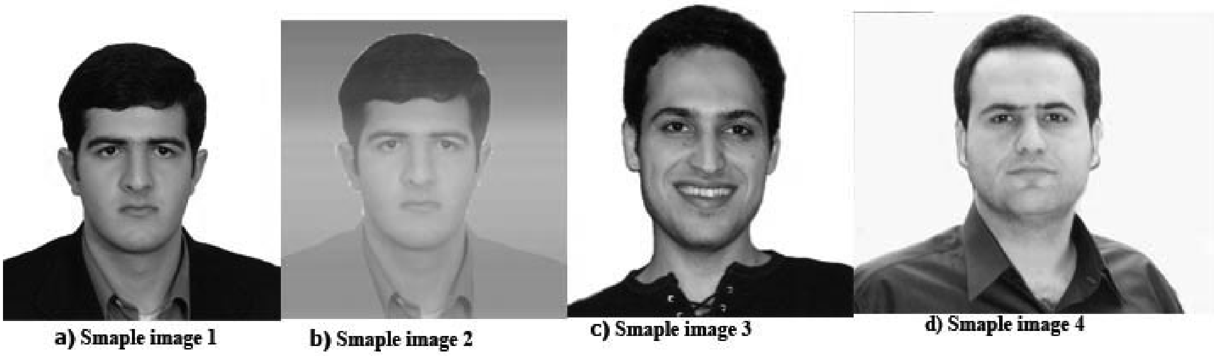

Figure 10. Sample images

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

Sample Image 1 compared with	similarity based on symbol	Ultimate similarity
Sample Image 2	98%	85.5%
Sample Image 3	42.8%	48.3%
Sample Image 4	51%	47.3%

We compared our new method with the expressed method in (7) which calculates the amount of similarity only based on symbol. These comparisons are shown in **figure 11**. As you see, it is clear that our new method is similar to 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, the efficiency of our new method is 12 percent more in average.

The time complexity of our new method is O(Log N), it is similar to the method which is only based on symbol (7), Because when we implemented this method, we used B^+ Tree like (3) for searching based on K_s . Of course the access time and search time can be optimized by multi dimensional B^+ Tree and R Tree structures.

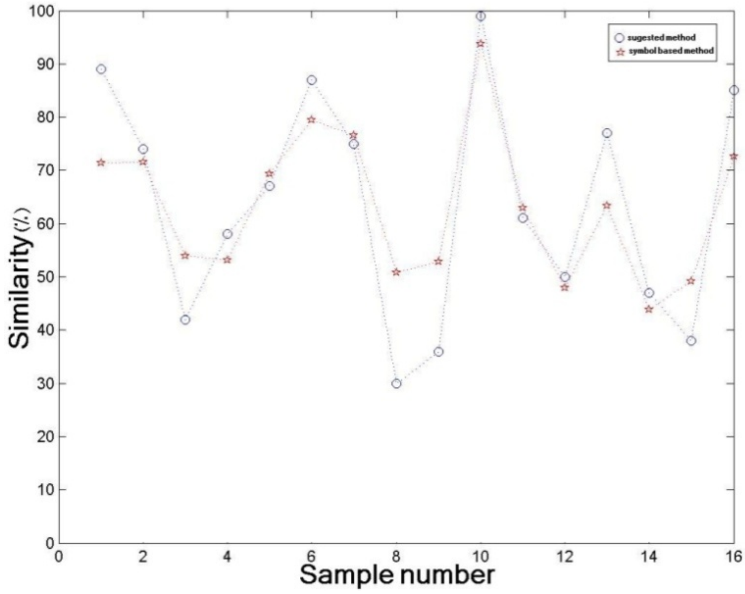

Figure 11. Comparison between the new method and the method in (2) for determining the similarity amount of images

CONCLUSION

After studying the tests and experiments, we reached the empirical conclusions as follows:

- This proposed method is an efficient method. In 200 sample tested images it can determine the amount of images similarity compared with the amount of real similarity with more than 97.5 percent accuracy.
- The efficiency of this proposed method is 12 percent more than the method which is only based on symbol to determine the amount of similarity.

Instead of using assurance ranges for matching images, we can use the fuzzy variables of colored spectrum histograms to determine the similarity amounts based on fuzzy rules. Furthermore, we can use symbol features and histograms of colored spectrums in addition to the texture features of images to determine the amount of similarity.

ACKNOWLEDGMENTS

This paper has been extracted from a research plan named “image retrieval based on symbol, color and texture for face recognition.” It was financed by the research department of Islamic Azad University, Jouybar Branch. I would appreciate the mentioned university for its support.

REFERENCES

1. S.UMBAUGH, “Computer Vision & Image Processing”, Indianapolis : Books craft, 1998.
2. R.R. Venkateswara and Subramanyam V.P. Rallabandi, “Rotation-invariant texture retrieval using wavelet-based hidden Markov Trees”, *Signal Processing,* Vol. 88, pp. 2593– 2598, 2008.
3. P. Punithaa, and D.S. Gurub, “Symbolic image indexing and retrieval by spatial similarity: An approach based on B-Tree”, *Pattern Recognition,* Vol. 41, pp. 2068 – 2085, 2008.
4. M.H. Shu and C.H. Chiun, “Retrieval of images by spatial and object similarities”, *Information Processing and Management*, Vol. 44, pp. 1214–1233, 2008.
5. V. P. Subramanyam and S.K. Sett, “Image retrieval system using R-Tree self-organizing map”, *Data & Knowledge Engineering*, Vol. 61, pp. 524–539, 2007.
6. R. Gonzalez and R. E. Woods, “Digital Image Processing, 3rd edition”, Addison Wesley publishing company, New York, USA, 2007.
7. F.L. George, “Expert Systems”, Macmillan Publishing Company, New Jersey, 1994.