Indian Journal of Science and Technology Vol. 3 No. 10 (Oct 2010) ISSN: 0974- 6846

A new method in determining the degree of similarity using symbols and average color spectrum characteristics

Hossein Shirgahi and Najmeh Danesh[1]

Islamic Azad University, Jouybar Branch, Jouybar, Iran; [1] *Islamic Azad University, Sari Branch, Sari, Iran, h.shirgahi@jouybariau.ac.ir, danesh@sariiau.ac.ir*

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 the appearance features. In this research, 250 samples of different people's faces have been taken in RGB mode as the input data and every image is changed to HSV colored mode. Then we calculate each image's symbols and the average colored of each symbol. Then we calculate the angles between applied symbols. In the next step we calculate the considered parameters based on symbols, average colored spectrum of symbols and the angles between them and they have been saved in the database. Finally, we determine the degree of similarity between images using features such as symbols, average colored spectrum of symbols and the angle between them. They were grouped as set. Based on the observed experimental results, this method's efficiency is 10 percent more than the method which is based on the symbol. Also it is more efficient 10.5 percent than the methods based on spatial and objective similarity or only based on the color.

Keywords: Image processing, the average spectrum of color, HSV color mode, symbolic images.

Introduction

Image processing is one of the most widely used scopes in computer. 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 aspects of image processing are image retrieval, image improvement and image compression. Image retrieval is the most widely used contexts as analyses the images to reach some scientific, economical and security goals (Umbaugh,1998).

In most applications, to do image retrieval economically for real environments, there are two important factors: running speed and accuracy. These are important in face recognition of image retrieval too. Out of the two, accuracy is assumed greater importance because of image retrieval for face recognition is used in military and high secure environments. In previous experiments' accuracy was low because only one or two features of an image was taken into consideration.

Venkateswara and Subramanyam (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 using the effect of textures which are extended in to sub groups' wavelets' coefficient. Experiment results show that this method reaches more punctual answers rather than previous methods which are based on wavelet in image texture retrieval.

In a research done by Punitha and Gurub (2008), indexing on symbolic images 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 (Punithaa & Gurub, 2008) (Fig.1).

Shu Ming Hsieh and Chiun Chieh (2008) perform the retrieval images based on spatial and objective similarities. They do it by object's conceptual information that is achieved by the images and spatial patterns. 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 (Shu Ming & Chiun Chieh, 2008).

Subramanyam and Sett (2007) have done an image retrieval system using R trees, They described the system based on visual image such as color, texture and design forms of descriptive files developed on the indexing information using R trees.

In the next section we present a fuzzy Histogram colors for color and to obtain recovery of incorrectly describing all forms of images. Experiments and tests on the Corel image database system did that performance for different images in the database and it was about 98 percent (Subramanyam & Sett, 2007).

In this study, the average spectrum of colored symbols for their images decide the accuracy of this method compared to previous work done. Often, single feature of the image for decision making was considered. This helps to resolve the problem existing between the

Indian Journal of Science and Technology Vol. 3 No. 9 (Sep 2010) ISSN: 0974- 6846

methods for extraction and analysis of two parameters-symbol and color images,

Determination of similarity based on symbol

In this research, first we calculate each image's symbols 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 the two images are the same, it means they are the same as color, shape, area and of course the sequence of three dimensional relations.

We use TSR model to determine the images' similarity based on symbols (Punithaa & Gurub, 2008). In this model the relations between the components of symbolic images is expressed by a Quadruple (L_a, L_b, L_c, θ). L_a, L_b and L_c are three symbols and θ is the angle between them. We construct these quadruples for each three symbols of the image. However, according to Fig.2, it should be noted that for each triple, such as A, B, C that are located in one line, we can offer six quadruples. We should select a case which has one of following conditions to consider only one of the cases equally for every three symbols in the image:

1 - $L_{i_1}, L_{i_2}, L_{i_3}$ labels 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

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

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

and $L_{i_1} = L_{i_2} = L_{i_3}$ where M is calculated according to equation 1.

$$M = Max(Dist(comp(L_{i_1}), comp(L_{i_3})), Dist(comp(L_{i_2}), comp(L_{i_3})) \quad (4)$$

Dist is a function to express the Euclidean distance between the middle points of each symbol. The function comp shows the center of considered symbol and Θ is the smallest angle between the middle points of components. it is defined according to equation 2.

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

Of course, if the three symbols are on one line, the angle will be calculated from equation 3.

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

We use equation 4 to display a unique number for quadruple (La, Lb, Lc, θ), Then we calculate K_S values for all created quaternaries of that image's symbols for each image. Then we save them in the database.

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

D_Θ is the number of parts that is considered for the angle. in this study we consider the angel in the range of [0..90] that each degree is one part. C_Θ is the number of part that the Θ belongs to (Punithaa & Gurub, 2008).

Finally, to calculate the amount of similarity of an image in the query (QI) with an image of the database (DBI), we perform as below:

first we compute all K_{SS} of the image according to equation 4, then we calculate the number of K_{SS} of the database image which is similar to it. The amount of similarity based on the symbol is calculated according to equation 5.

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

In this equation, $TNS_{K_S}(DBI)$ is a number of KSs of a DBI that are the same as K_{SS} of QI. $TN_{K_S}(QI)$ is the total number of K_{SS} of QI.

Determination of similarity based on the average color

In this study one of the parameters is used for the determination of the similarity of images is color. We use the average colored to determine this feature. The average colored is the simplest presentation of images colored vector feature (Gonzalez & Woods, 2007).

According to the researches, HSV colored space coordinates to humans understanding of color. At the first phase, the input image received in RGB colored mode is converted to HSV colored mode. Then we quantize this space to 36 separate parts. We used 4 different values for H and 3 different values from V and S. So we consider a number from 1 to 36 based on the average colored spectrum for each image. Fig.3 shows a sample image with images of spectrums H, S and V. The value obtained based on the average colored spectrum H, S and V is one of the comparable parameters in images similarity.

Determination of the amount of similarity based on combined features as symbol and the average colored spectrums

After expressing the amount of similarity based on the symbol (SS) and expressing the amount of similarity of average colored spectrums, we improve the method which determines the amount of similarity based on symbol. We add color feature to each symbol. Finally we calculate the final amount of similarity (ST) based on features of the symbols, average colored spectrums of symbols and the angle between them. To achieve this goal, we use equation 5 instead of equation 4 to obtain unique values of K_S.

$$K_S = D_\theta M_{C_1}(L_{i_1} - 1)m^2 + D_\theta M_{C_2}(L_{i_2} - 1)m + D_\theta M_{C_3}(L_{i_3} - 1) + (C_\theta - 1) \quad (6)$$

Indian Journal of Science and Technology Vol. 3 No. 9 (Sep 2010) ISSN: 0974- 6846

M_{C1}, M_{C2} and M_{C3} are the values obtained based on the average colored spectrums H, S and V for each symbol L_{i1}, L_{i2} and L_{i3}. In this section, we do not consider average colored spectrum for the whole image but we consider average colored spectrums H, S and V for each symbols of the used image. Then we calculate the unique values of K_S according to equation 6. This equation is based on symbols, the average colored spectrums of symbols and the angles between them.

Fig.1. An example of symbolic images (Punithaa & Gurub, 2008)

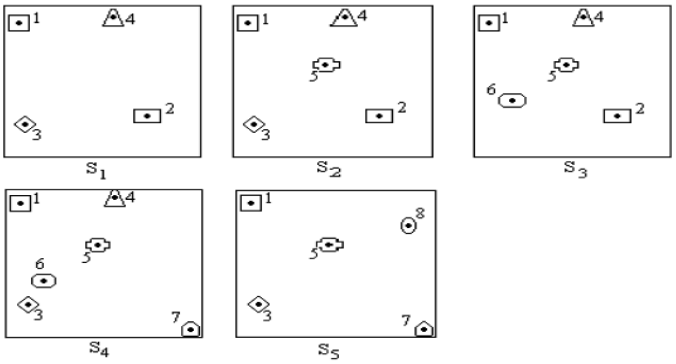


Fig. 2.Triangular spatial relationship (Punithaa & Gurub, 2008)

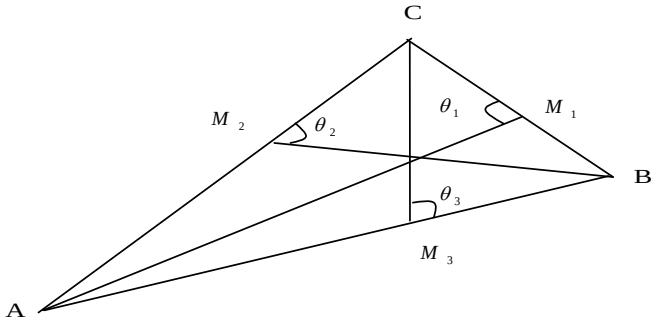
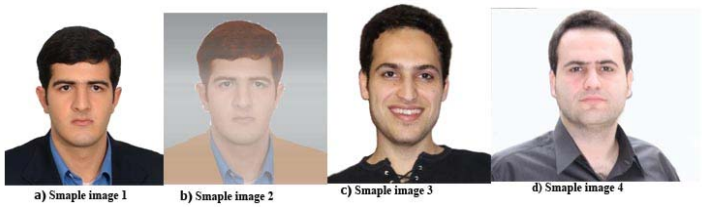


Fig.3.The color spectrum H, S and V in sample image

Fig. 4. Sample images



Finally, we perform as below to determine the amount of similarity of an image in the query (QI) with an image in the database (DBI): first we compute all K_S of the query image according to equation 6 and then we calculate the number of K_S of an image database which is the same as that one. So the amount of similarity is calculated according to equation 5 was explained in Section 2.

Implementations and tests

We implemented this research by MATLAB 7 software. First we saved 250 images of different people in the designed database. We investigate SQL Server 2000 as the system database. We stored some information in it such as: images physical addresses, symbols of each image with the symbol number, image number and obtained K_{SS} for each image.

Then in order to compare a sample image with one or more other images, we investigate that if it is a new image or it is not available in the database. We compute K_{SS} parameters based on equation 6. But if it is not a new image, we retrieve respective parameters of this image from the database. Then we calculate the ultimate amount of similarity based on equation 5.

Fig.4 shows a sample image and Table 1 shows the comparison between the amount of similarity of image 1 with sample Images 2, 3 and 4.

We compared our new method with the expressed method (Punithaa & Gurub, 2008) which calculates the amount of similarity only based on symbol. Also we compared it with the method expressed by Shu Ming & Chiun Chieh (2008) which calculates the amount of similarity based on spatial and objective.

As shown in Fig.5, it is clear that our new method is similar to those based only on symbol applicable to images that color does not have such an effect. But for those images where color has an important effect on the similarity, the efficiency of our new method is 8 to 12 percent more, in an average. Also the efficiency of the expressed method is 8 to 13 percent more than the method which determines the amount of similarity based on spatial and objective similarities and it is nearer to the real cases.

The time complexity of our new method is O(Log N) (Punithaa & Gurub, 2008), it is similar to the method which is only based on symbol because when we implemented this method, we used B^+ tree like (Shu Ming & Chiun Chieh, 2008) for searching based on K_{SS}. Of course, the access time and search time can be optimized by multi dimensional B^+ tree and R tree structures.

Indian Journal of Science and Technology Vol. 3 No. 9 (Sep 2010) ISSN: 0974- 6846

Table 1. Comparison of similarity 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%	83%
Sample Image 3	42.8%	46%
Sample Image 4	51%	48%

Conclusion

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

This proposed method is an efficient method. In 250 sample tested images, it can determine the amount of images similarity with more than 90 percent accuracy compared to the real similarity.

The efficiency of this proposed method is 10 percent more than the method based only on symbol to determine the amount of similarity.

Also, the efficiency of this proposed method is 10.5 percent more than the method which is based on spatial and objective similarities and the obtained answers are nearer to real similarity amounts.

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.

Acknowledgements

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. Gonzalez R, Woods RE (2007) Digital image processing. 3rd ed, Addison Wesley Publ. Co., NY.
2. Punithaa P, Gurub DS (2008) Symbolic image indexing and retrieval by spatial similarity:An approach based on B-tree. Pattern Recognition. 41, 2068–2085.
3. Shu Ming H and Chiun Chieh H (2008) Retrieval of images by spatial and object similarities. *Information Processing & Management*. 44, 1214–1233.
4. Subramanyam VP and Sett SK (2007) Image retrieval system using R-tree self-organizing map. Data & Knowledge Engg. 61, 524–539.
5. Umbaugh S (1998) Computer vision & image processing. Indianapolis: Books craft.
6. Venkateswara RR and Subramanyam VP (2008) Rotation-invariant texture retrieval using wavelet-based hidden Markov trees. *Signal Processing*. 88, 2593–2598.