Study of Face Recognition Approach Based on Similarity Measures

Mohamed A. El-Sayed[1] and Noha Aboelwafa[2]

[1] Mathematics department, Faculty of Science, Fayoum University,63514, Fayoum , Egypt
Assistant professor, Dept of CS, CIT College, Taif University,21974,Taif, KSA

[2] Mathematics Department, Faculty of Science, South Valley University,
Qena, Egypt

Abstract

Detecting the similarity of face image aims to determine the image of a face for verification purpose of documents such as passport, driving license, ID cards, etc. Similarity measures are essential to solve many pattern recognition problems such as classification, clustering, and retrieval problems. There are efforts in finding the appropriate measures among such a plethora of choices because it is of fundamental importance to solve our problems. A new approach for face recognition based on similarity measure method is introduced. In addition, We apply various measure classes to increase the efficiency of the proposed method. Experimental results show that these similarity measures can give an useful way for measuring the similarity between fuzzy sets.

Keywords: *Algorithms, Image processing, Fuzzy sets, Similarity measures, face recognitions.*

1. Introduction

The text must be in English.

Measuring the similarity between objects plays an important role in many fields of computer science, such as image processing, image retrieval, image compression, pattern recognition, clustering, and information retrieval problems, etc. Objective measures or measures of comparison are required to test the performance of applying algorithms to an image, to compare the output image. Visual tasks are often based on the evaluation of similarities between image-objects represented in an appropriate feature space. The performance of content-based query systems depends on the definition of a suitable similarity measure [1][2].

The interest in the development of content-based image retrieval (CBIR) system is increasing because of the growth in the number of image databases in many domains such as multimedia libraries, medical images and geographical information systems. In CBIR systems, the comparison of two images is a fundamental operation and is rarely made based on exact match. An image can be represented by a feature vector, where each element is associated to an attribute (or feature) of a image. These attributes are represented, in general, by single numerical values obtained by feature extractors. The similarity of two images is obtained by computing the similarity (or dissimilarity) between their feature vectors [3].

Several measures have been proposed to measure the similarity between fuzzy sets or images [4-11].

There is no generic method for selecting a suitable similarity measure or a distance measure. However, a prior information and statistics of features can be used in selection or to establish a new measure. Van der Weken et al. [12] gave an overview of similarity measures, originally introduced to express the degree of comparison between fuzzy sets, which can be applied to images. These similarity measures are all pixel-based, and have therefore not always satisfactory results. To cope with this drawback, in [13] they proposed similarity measures based on neighborhoods, so that the relevant structures of the images are observed better. The authors in [14] reviewed some existing similarity measures, showed that these measures are not always effective in some cases and illustrated the problem in the context of colorectal cancer diagnosis by similarity measure between fuzzy rough sets. There are several similarity measures that are proposed and used for varied purposes, see [15-24].

Many method was described and applied using entropy types to handle the face recognition and edge detection problems.[25-29].

From the scientific and mathematical point of view, *distance* is defined as a quantitative degree of how far apart two objects are. Synonyms for *distance* include dissimilarity. Those distance measures satisfying the metric properties are simply called *metric* while other non-metric distance measures are occasionally called *divergence*.

Synonyms for *similarity* include proximity and similarity measures are often called *similarity coefficients*. A distance measure and a similarity measure are denoted as d_x and s_x, respectively throughout the rest of the paper.
In this paper, we propose a novel approach to derive the similarity between two images using new similarity measure, by representing each numerical value of their feature vectors as a fuzzy set, instead of a single value. This representation takes into account the uncertainty presents in the extraction process of features and consequently, increases the precision rate in the image retrieval process. In order to test our new approach, we used ORL face database and various similarity measures of L_p metrics for $p \in \{1,2,3,4\}$ and a proposed similarity measures. The results obtained by the proposed approach present higher performance than the traditional ones.

The rest of this paper is organized as follows: Section 2 presents the mathematical foundations of fuzzy sets and digital images. Sections 3 describes the proposed similarity measures and investigating its properties. Experimental results on real data are outlined in section 5, and finally, the conclusions are given in section 6.

2. Similarity Measures

Similarity measure is defined as the distance between various data points. The performance of many algorithms depends upon selecting a good distance function over input data set. While, similarity is a amount that reflects the strength of relationship between two data items, dissimilarity deals with the measurement of divergence between two data items [30-32].
The theory of fuzzy sets $F(X)$ was proposed by Zadeh [33,5]. A fuzzy set A in a universe X=$\{x_1, x_2, ... x_n\}$ is characterized by a mapping $\chi_A : X \to [0,1]$, which associates with every element x in X a degree of membership $\chi_A(x)$ of x in the fuzzy set A. In the following, let a= $\{a_1, a_2, ... a_n\}$ and b=$\{b_1, b_2, ... b_n\}$ be the vector representation of the fuzzy sets A and B respectively, where a_i and b_i are membership values $\chi_A(x_i)$ and $\chi_B(x_j)$ with respect to x_i and x_j $(i, j = 0,1,2\ldots,n)$ respectively. Furthermore, suppose F(X) be the class of all fuzzy sets of X, $A^c \in F(X)$ is the complement of $A \in F(X)$.
There is no unique definition for the similarity measure, but the most common used definition is the following [34-36] .

Definition 2.1
A mapping S: $FS(X) \times FS(X) \to [0,1]$, is said to be measures between Fuzzy Sets $A \in FS(X), B \in FS(X)$ if $S(A,B)$ satisfies the following properties:

$(SP_1): 0 \le S(A,B) \le 1$, for all $A,B \in F(x)$
$(SP_2): S(A,B) = 1$, iff $A = B$
$(SP_3): S(A,B) = S(B,A)$, $A,B \in FS(X)$
(SP_4): If $A \subseteq B \subseteq C$ for all $A,B,C \in F(x)$ then $S(A,B) \ge S(A,C)$ and $S(B,C) \ge S(A,C)$.

Definition 2.2
If $S(A,B)$ is similarity measure defined as above, then

$$d(A,B) = 1 - S(A,B) \tag{1}$$

is a distance measure between A and B.

Based on this definition several similarity measures have been proposed [4,10]. The first similarity measure is based on the fuzzy Minkowski distance d_r, and the observation that the smaller the distance between A, and B, the greater the similarity between A, and B. This observation leads to the following similarity measure $S_1(A,B)$:

$$S_1(A,B) = 1 - \left[\frac{1}{n} \sum_{i=1}^{n} | a_i - b_i |^r \right]^{1/r}, \quad r \ge 1. \tag{2}$$

There are other similarity measures which are also based on a distance such as $S_2(A,B)$ and $S_3(A,B)$.

$$S_2(A,B) = 1 - \frac{\sum_{i=1}^{n} |a_i - b_i|}{\sum_{i=1}^{n} (a_i + b_i)} \tag{3}$$

and

$$S_3(A,B) = \frac{2^r \sum_{i=1}^{n} (a_i \cdot b_i)}{\sum_{i=1}^{n} a_i^2 + \sum_{i=1}^{n} b_i^2 + (2^r - 2) \sum_{i=1}^{n} (a_i \cdot b_i)}, r \ge 0 \tag{4}$$

The set-theoretic similarity measures are the most suitable for measuring similarity between overlapping fuzzy sets. Matching function-based similarity measures are existed such as:

$$S_4(A,B) = \frac{\sum_{i=1}^{n} (a_i \cdot b_i)}{\max(\sum_{i=1}^{n} a_i^2 , \sum_{i=1}^{n} b_i^2)} \tag{5}$$

The larger the value of the above all similarity measures, the more the similarity between the fuzzy sets A and B. But, with all this number of existed similarity measures, there is not any relation between them. In other words, there are not defined operations that control the behavior of these measures when they are applied concurrently.

3. Face Recognition Algorithm

In this section, we will describe a face recognition algorithm using the neighborhood feature vectors with various similarity measures . Automatic face recognition systems try to find the identity of a given face image according to their memory. The memory of a face recognizer is generally simulated by a training set. In this paper, our training set consists of features extracted from known face images of different persons. Thus, the task of the face recognizer is finding most similar feature vector among the training set to the feature vector for a given test image. Here, we want to recognize the identity of a person where his image is given to the system as a test image.
Let $f(x, y)$ be the gray value of the pixel located at the point (x, y). In a digital image of size $M\times N$ (M rows, N columns), $1\leq x \leq M$ and $1\leq y \leq N$. Also, let f_1, f_2, …., f_n are face images in available database.

In the training phase, we extract feature vectors for each image in the training set. Let v_k be a training image vector of person k which has a pixel resolution of $M\times N$. The length of the vector v_k will be $M\times N$. Let g_k (x, y) be the average of the neighborhood of the value located at the point (x, y). The average gray value for the 3×3 neighborhood of each pixel is calculated as:

$$g_k(x,y)=\frac{1}{9}\sum_{a=-1}^{1}\sum_{b=-1}^{1} f_k(x+a,y+b) \qquad (6)$$

while computing the average gray value, disregard the two rows from the top and bottom and two columns from the sides.
In the recognition phase (or testing phase), we give a test image j of a known person. As in the training phase, we compute the feature vector of this person using the neighborhood pixels and obtain v_j . In order to identify v_j ,we compute the similarities between v_j and all of the feature vectors v_i's in the training set. The similarity between feature vectors can be computed using classic metric. The identity of the most similar will act the output of the proposed face recognizer. It stores the labels of images which have the largest similarity values of the face database, and also values will order in a descending way.

The proposed algorithm of face recognition based on similarity measure classes consists of the following steps: (See Figure 1)

Step1: Read the gray values of face image f_i, where $0\leq f_i(x,y) \leq 255$, $i=1,2,\ldots,n$.

Step2: Normalize the gray values of face image f_i $(x, y) \in [0,1]$.

Step3: Compute the average value of the neighborhood g_i (x, y) of each value f_i (x, y).

Step4: To simplify the calculations ,for each face image, we save g_i (x, y) as vector v_i , $i=1,2,\ldots,n$. Now we have feature space $V=\{v_1,v_2,\ldots,v_n\}$ by repeating of steps(1-4). These vectors are corresponding to the face images in the database.

Step5: Randomly, select one face image from the database, say f_k, $1 \leq k \leq n$. Then the vector $A= v_k$ is corresponding to the tested face image.

Step6: For all i, $1 \leq i \leq n$. Calculate the similarity value $S(A,B)$ where $B= v_i$, and store the label i and its $S(A,B)$.

Step7: Get the index of the largest similarity value ,and show its face image.

Step8: Stop.

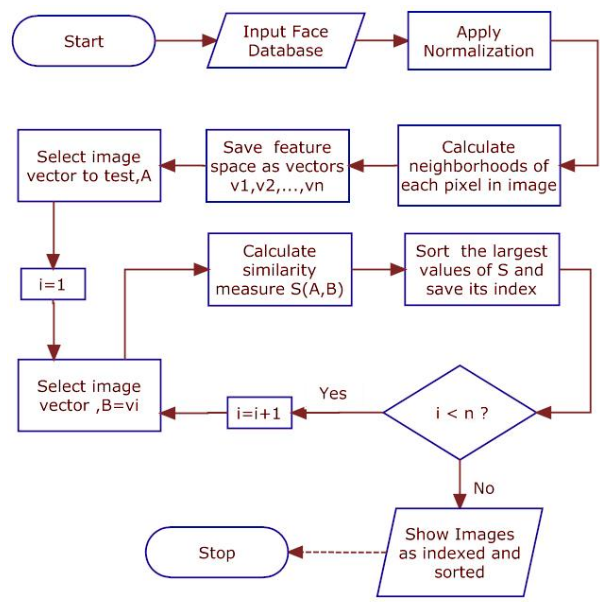

Fig. 1 Flowchart that illustrate the steps of proposed algorithm for face recognition in image sets.

4. Results and Discussion

The previous algorithm is based on the definitions and the theorem given in section 3. The current section is divided into two subsections. The first subsection explains how the proposed algorithm run by a given example. The next subsection shows the experimental results with real database, ORL. It present comparative results applied many types of similarity measure classes to determine the best similarity measure class.

4.1 Simple data set

We will explain the proposed algorithm by using simple face images. For example, let the face data base contains ten face images of various persons, $P1$, $P2$, $P3$, ..., $P10$. The diagram of the face recognition system that will be implemented is shown in Figure 2. Each person in training phase has a special feature vector generated by feature extractor, $B1$, $B2$, $B3$, ..., $B10$. In the other hand, let A is a feature vector generated by the feature extractor of the test image, here let $A=v_4$. Applying one similarity measure from the measures in equations (2), , S_1 with r=1.

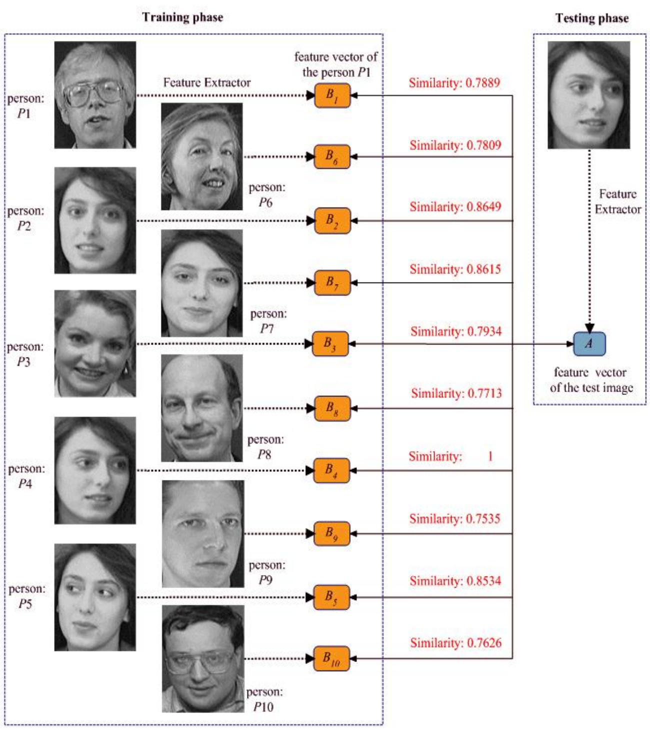

Fig.2 Diagram of a face recognizer in the proposed algorithm.

Table 1: The similarity values of face database and $A=v_4$

Image	***P1***	***P2***	***P3***	***P4***	***P5***
S(A,B)	0.7889	0.8649	0.7934	1	0.8534
Image	***P6***	***P7***	***P8***	***P9***	***P10***
S(A,B)	0.7809	0.8615	0.7713	0.7535	0.7626

The similarity values of the persons $P1$, $P2$, $P3$, ..., $P10$ are shown in Table 1. In this table, the largest similarity value is pound between the test image and the image of the person $P4$. According to this value is equal to one , so the test image and the image of $P4$ are the same. In this way, the images of persons, $P2$, $P7$, and $P5$ have the similarity values, 0.8649 , 0.8615, and 0.8534, respectively. i.e. $P2$, $P7$, and $P5$ are the largest similarities of the test image, respectively. In other hand, the image of person $P9$ has the smallest similarity of the test image. In this case, the out puts of the system are shown in Figure 3.

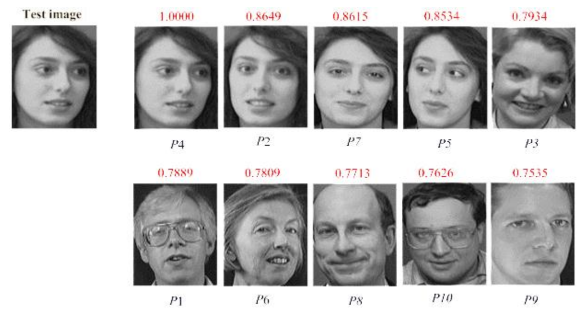

Fig.3 The results of the proposed algorithm on limited data set.

4.2 Experimental results on ORL database

There is a well-known face database which can be downloadable from the AT&T Laboratories, Cambridge at http://www.uk.research.att.com/facedatabase.html. or Cambridge University site at http://www.cl.cam.ac.uk/ Research/DTG/attarchive/pub/data/. The ORL database contains 400 face images from 40 individuals in different states.

The total number of images for each person is 10. None of the 10 samples is identical to any other sample. They vary in position, rotation, scale and expression. Little variation of illumination, slightly different facial expressions and details are present in the face images. The changes in orientation have been accomplished by rotating the person a maximum of 20 degree. For the same subject; each person has also changed his/her facial expression in each of 10 samples (open/close eye, smiling/not smiling). The changes in scale have been achieved by changing the distance between the person and the video camera. For some individuals, the images were taken at different times, varying facial details (glasses/no glasses). Each image was digitized and presented by a 92×112 pixel array whose gray levels ranged between 0 and 255.See Figure 4.

To test all images in ORL face database, we run the proposed algorithm 400 times. each time with several similarity measure such as S1,S2,S3,S4.

Each individual (or test image) in data base has ten face images in different states. Table 2, shows the test image number and the corresponding number of right similarities detected from the ten of actual face images in special measure class. For example, the test image of number 1 , the system detected 3 similarities images with S1, 4 similarities image with S2 and S3 , and 3 similarities images with the measure S4.

Fig.4 Sample of face images from ORL data set for a particular subject.

Table 2: Right similarities numbers in ORL face recognized

Test Image No	Numbers of right similarities images detected			
	s1	s2	S3	S4
1	3	4	4	3
2	7	7	3	4
3	6	7	3	4
4	7	7	6	6
5	7	7	7	6
6	6	6	3	4
7	7	9	7	4
8	8	9	7	6
9	8	8	6	6
10	4	4	3	4
11	10	10	9	10
12	10	10	10	10
13	10	10	10	10
14	10	10	10	8
15	10	10	9	10
16	10	10	9	9
…	…	…	…	…
400	4	3	3	4
Average	7.2	7.1	6.7	5.3

Figure 5 summaries the results of Table 2. X-axis represents the different types of the similarities measures classes . Y-axis represents the average value of detected right images of the same individual through 400 images with each type of measures. We note that, the performance of $S_i(A,B)$ is the same as $S_i(A,B)$, i=1,2,3,4. Also, the similarity measure class S_1 is the best performance, then S_1, S_2, and S_3, respectively, But the measure S_4 is the less efficient.

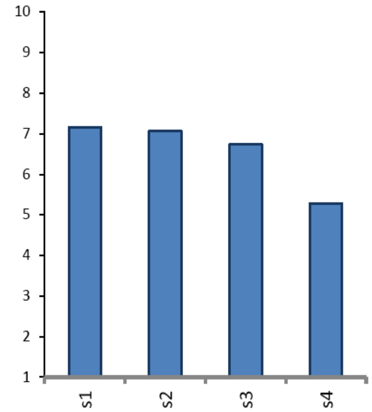

Fig.5 The performance of $Si(A,B)$.

Table 3, shows how many numbers of test images which are used with S_i of detected $1 \le n = number \le 10$ face images of the same individual. For example, In case S1, there are 67 images, each of them is used as a test image. The used similarity here succeed to detect all images(10) of the same individual. Also, there are 65 images succeed to detect 9 images of the same individual of the test image. Figure 6 summaries the results in Table 3. The chart satisfied the previous that the best performance of measures classes is S_1.

Table 3: Retrieval Images Numbers

Similarity Measure Class	Images number of the same person										
	1	2	3	4	5	6	7	8	9	10	sum
s1	0	5	6	37	59	46	61	54	65	67	400
s2	0	5	10	34	63	53	55	54	61	65	400
s3	2	2	22	46	56	51	63	62	46	50	400
s4	4	23	41	69	105	54	54	27	8	15	400

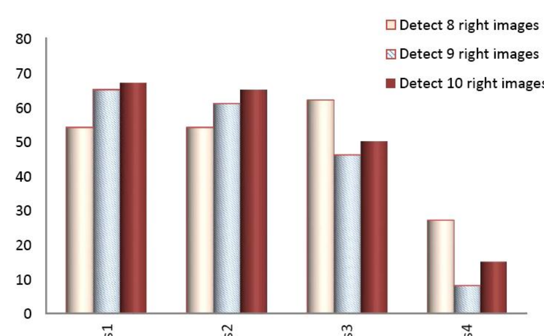

Fig.6 Number of the test images which used to detect individual images.

5. Conclusions

This paper has described an efficient method for face recognition based on different types of similarity measures of fuzzy sets. The proposed method has been implemented and tested on the ORL database with the different classic similarity measures. The experimental results are explained. The similarity measure based on the fuzzy Minkowski distance d_r get better evaluation of the similarity between fuzzy sets. The measure, S_1 with r=1, gives high performance than the studied classical similarity measures.

References

[1] Sung-Hyuk Cha, Comprehensive Survey on Distance/Similarity Measures between Probability Density, International Journal Of Mathematical Models And Methods In Applied Sciences, Issue 4, Volume 1, pp.300-307, 2007.

[2] Mohamed A. El-Sayed, A New Algorithm Based Entropic Threshold for Edge Detection in Images. IJCSI International Journal Of Computer Science Issues, 8(5)(1) 71-78, 2011.

[3] Hjaltason Gisli R., Hanan Samet, Index-driven similarity search in metric spaces (Survey Article). ACM Trans. Database Syst., 2003.

[4] S. H. Kwon, A similarity measure of fuzzy sets, Journal of Fuzzy Logic and Intelligent Systems, vol. 11, No.3, pp. 270-274, Korea Fuzzy Logic and Intelligent Systems Society, 2001.

[5] X. Liu, Entropy, distance measure and similarity measure of fuzzy sets and their relations, Fuzzy Sets and Systems, vol. 52, pp. 305-318, 1992.

[6] X. Wang, B.De Baets, E.E.Kerre, A comparative study of similarity measures, Fuzzy Sets and Systems, vol. 73, pp. 259-268, 1995.

[7] S. M. Chen, Measures of similarity between vague sets, Fuzzy Sets and Systems, vol. 74, pp. 217-223, 1995.

[8] W.J. Wang, New similarity measure on fuzzy sets and on elements, Fuzzy Sets and Systems, vol. 85, pp. 305-309, 1997.

[9] V. Di Gesu, V. Starovoito, Distance-based functions for image comparison, Pattern Recognition Letters, vol. 20, pp. 207-214,1999.

[10] Monev V., Introduction to Similarity Searching in Chemistry, MATCH Commun. Math. Comput. Chem. 51 pp. 7-38 , 2004.

[11] Gavin D.G., Oswald W.W., Wahl, E.R., and Williams J.W., A statistical approach to evaluating distance metrics and analog assignments for pollen records, Quaternary Research 60, pp 356–367, 2003.

[12] D. Van der Weken, M. Nachtegael , E.E. Kerre, An overview of similarity measures for images, Proceeding of ICASSP 2002 (IEEE International Conference on Acoustics, Speech and Signal processing), Orland ,US, pp.3317-3320, 2002.

[13] D. Van der Weken, M. Nachtegael , E.E. Kerre, Using similarity measures and homogeneity for the comparison of images, Image and Vision Computing, vol. 22, pp.695-702, 2004.

[14] C. Zhang, H. Fu, Similarity measures on three kinds of fuzzy sets, Pattern Recognition Letters, vol. . 27, no. 12, pp. 1307-1317, 2006.

[15] Ismat Beg, Samina Ashraf , Similarity Measures For Fuzzy Sets, Appl. and Comput. Math., V.8, N.2, pp.192-202, 2009.

[16] Zwick, R., Carlstein, E., Budesco, D.V. Measures of similarity amongst fuzzy concepts: A comparative analysis, Int. J. Approximate Reasoning, V.1, pp.221-242, 1987.

[17] Ashby, F.G. Towards a unified theory of similarity and recognition, Psychological Review, V.95, N.1, pp.124-150, 1988.

[18] Jain, R., Jayaram, Murthy, S.N., Chen, P.L-J., Chatterjee, S. Similarity measures for image databases, IEEE Int. Conference on Fuzzy Systems, V.3, pp.1247-1254, 1995.

[19] Santini, S., Jain, R. Similarity is a geometer, Multimedia Tools and Applications, V.5, N.3, pp.277-306, 1997.

[20] Santini, S., Jain, R. Similarity measures, IEEE Transactions on Pattern Analysis and Machine Intelligence, V.21, N.9, pp.871-883, 1999.

[21] Williams, J., Steele, N. Diference, distance and similarity as a basis for fuzzy decision support based on prototypical decision classes, Fuzzy Sets and Systems, V.131, pp.35-46, 2002.

[22] Cross, V.V., Sudkampm, T.A Similarity and Compatibility in Fuzzy Set Theory, Physica-Verlag Heidelberg, 2002.

[23] Wang, X., Kerre, E.E. Reasonable properties for the ordering of fuzzy quantities (i), Fuzzy Sets and Systems, V.118, pp. 375-387, 2001.

[24] Wang, X., Kerre, E.E. Reasonable properties for the ordering of fuzzy quantities (ii), Fuzzy Sets and Systems, V.118, pp.387-405, 2001.

[25] YuDong Zhang and LeNan Wu, Pattern Recognition via PCNN and Tsallis Entropy, Sensors journal, 8, 7518-7529, 2008.

[26] B.H. Shekar, M. Sharmila Kumari, Leonid M. Mestetskiy, Natalia F. Dyshkant, Face recognition using kernel entropy component analysis, Neurocomputing journal, Volume 74, Issue 6, 1053–1057, 2011

[27] Mohamed A. El-Sayed, S. Abdel-Khalek, Eman Abdel-Aziz, Study of Efficient Technique Based On 2D Tsallis Entropy For Image Thresholding. International Journal On Computer Science And Engineering (IJCSE), ,3,9, 2011.

[28] Mohamed A. El-Sayed, A New Algorithm Based Entropic Threshold for Edge Detection in Images. IJCSI International Journal Of Computer Science Issues, 8, 1, 2011.

[29] Mohamed A. El-Sayed, Tarek Abd-El Hafeez, New Edge Detection Technique based on the Shannon Entropy in Gray Level Images. International Journal on Computer Science and Engineering (IJCSE), 3,6,2011.

[30] Anil Kumar Patidar , Jitendra Agrawal , Nishchol Mishra, Analysis of Different Similarity Measure Functions and their Impacts on Shared Nearest Neighbor Clustering Approach , International Journal of Computer Applications ,Vol.40, No.16, 2012

[31] Anna Huang, “Similarity Measures for Text Document Clustering”, NZCSRSC 2008, April 2008, Christchurch, New Zealand.

[32] Kazem Taghva and Rushikesh Veni, “Effects of Similarity Metrics on Document Clustering”, 2010 Seventh International Conference on Information Technology.

[33] L. A. Zadeh, Fuzzy sets, Inform. Control, vol. 8, pp. 338-353, 1965.
[34] Bustince, H. Indicator of inclusion grade for interval valued fuzzy sets: Application to approximate reasoning based on interval-valued fuzzy sets, Int. J. Approx. Reasoning, V.23, pp.137-209, 2000.
[35] S. Omran and M. Hassaballah, A new class of similarity measures for fuzzy sets, International Journal of Fuzzy Logic and Intelligent Systems, vol. 6, no. 2, pp. 100-104, 2006.
[36] W. Wang, X. Xin, Distance measure between intuitionistic fuzzy sets, Pattern Recognition Letters, vol. 26, pp. 2063-2069, 2005.

.