International Journal of Fuzzy Logic and Intelligent Systems, vol. 6, no. 2, June 2006 pp. 100−104

A New Class of Similarity Measures for Fuzzy Sets

Saleh Omran and M. Hassaballah

Mathematics Department, Faculty of Science, South Valley University, Qena, Egypt.

Abstract

Fuzzy techniques can be applied in many domains of computer vision community. The definition of an adequate similarity measure for measuring the similarity between fuzzy sets is of great importance in the field of image processing, image retrieval and pattern recognition. This paper proposes a new class of the similarity measures. The properties, sensitivity and effectiveness of the proposed measures are investigated and tested on real data. Experimental results show that these similarity measures can provide a useful way for measuring the similarity between fuzzy sets.

Key words : Fuzzy sets, Similarity measures, Images comparison

1. Introduction

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, 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].

Several measures have been proposed to measure the similarity between fuzzy sets or images [2-7]. 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. [8] 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 [9] they proposed similarity measures based on neighbourhoods, so that the relevant structures of the images are observed better. In this way 13 similarity measures were found to be appropr·-te for the comparison of images.

Another type of similarity measures between intuitionistic fuzzy sets (IFSs) was proposed by Weiqiong Wang et al [10], some distance measures and the corresponding proofs are given, and the relations between similarity measure and distance measure of IFSs are analyzed. Measuring the degree of similarity between three fuzzy sets under unifying form and between IFSs is presented in [11]. The authors reviewed some existing similarity measure, 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.

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 4, and finally, the conclusions are given in section 5.

2. Mathematical foundations of fuzzy sets

2.1 Fuzzy sets and digital images

The theory of fuzzy sets *F(X)* was proposed by Zadeh [12]. A fuzzy set A in a universe X= $\{x_1, x_2, ...x_n\}$ is characterized by a mapping $\chi_A : X \rightarrow [0,1]$, 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 $(i, j = 0,1,2...n)$ respectively. Furthermore, suppose *F(X)* be the class of all fuzzy sets of X, $A^c \in F(X)$ is the complement of $A \in F(X)$.

In order to model the intersection or union between two fuzzy sets the $\wedge$ and $\vee$ operators will be used to refer to the minimum and maximum respectively. The cardinality of a finite crisp set is given by the number of elements in that set. This concept can be extend to fuzzy sets: the sigma count of a fuzzy set A (with finite support) in a universe X is defined as

Manuscript received May. 18, 2006; revised Jun. 10, 2006.

$$|A| = \sum_{x_i \in X} \chi_A(x_i)$$

A digital image can be identified with a fuzzy set that takes values on the grid points (x,y), with $x,y \in \mathbb{N}$, $0 \le x \le M$ and $0 \le y \le N$, ($M, N \in \mathbb{N}$). Consequently, for two digital images A and B, one have that A,B ∈ F(X), with $X = \{(x, y) | 0 \le x \le M, 0 \le y \le N\}$ a discrete set $\{n = 1, 2, ...MN\}$ of image points.

2.2 Similarity measures

There is no unique definition for the similarity measure, but the most common used definition is the following.

Definition 2.1. A similarity measure is a function assigning a similarity value to the pair of fuzzy sets (A,B) that indicates the degree to which A and B are equal or how similar they are. This function must be reflexive, symmetric and min-transitive. On other word, A mapping $S : F(X) x F(X) \rightarrow [0,1]$ is said to be a similarity measure between fuzzy sets $A \in F(X)$ and $B \in F(X)$, if $S(A, B)$ satisfies the following properties:

(SP1) $S(A,B) = S(B,A), \quad A,B \in F(X)$;

(SP2) $S(D, D^c) = 0$, if D is a crisp set ;

(SP3) $S(E,E) = \max_{A,B \in F(X)} S(A,B)$, for all $E \in F(X)$;

(SP4) If $A \subseteq B \subseteq C$ for all $A,B,C \in F(X)$ then $S(A,B) \ge S(A,C)$ and $S(B,C) \ge S(A,C)$.

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

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

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

$$S_2(A,B) = 1 - \frac{\sum_{i=1}^{n} |a_i - b_i|}{\sum_{i=1}^{n} (a_i + b_i)}, \qquad (2)$$

$$S_3(A,B) = 1 - d_\infty(a,b) = 1 - \max_i(|a_i - b_i|) \qquad (3)$$

Another type of similarity measures are based on the set-theoretic. These measures are based on the sigma count and the intersection or union of two fuzzy sets:

$$S_4(A,B) = \frac{|A \cap B|}{|A \cup B|} = \frac{\sum_{i=1}^{n} (a_i \wedge b_i)}{\sum_{i=1}^{n} (a_i \vee b_i)} \qquad (4)$$

$$S_5(A,B) = \frac{1}{n}\sum_{i=1}^{n} \frac{(a_i \wedge b_i)}{(a_i \vee b_i)} \qquad (5)$$

$$S_6(A,B) = \frac{|A \cap B|}{\min(|A|,|B|)} = \frac{\sum_{i=1}^{n} (a_i \wedge b_i)}{\min(\sum_{i=1}^{n} a_i, \sum_{i=1}^{n} b_i)} \qquad (6)$$

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

$$S_7(A,B) = \frac{\sum_{i=1}^{n} (a_i \cdot b_i)}{\max\left(\sum_{i=1}^{n} a_i^2, \sum_{i=1}^{n} b_i^2\right)} \qquad (7)$$

The larger the value of the above all similarity measures, the more the similarity between the fuzzy sets A and B. These similarity measures are mentioned in detail in [2,9].

The classical measures for images comparison are the mean square error (MSE),

$$MSE(A,B) = \frac{1}{n}\sum_{(x,y) \in X} |A(x,y) - B(x,y)|^2 \qquad (8)$$

and the peak-signal-to-noise-ratio (PSNR)

$$PSNR(A,B) = 20\log_{10}\left(\frac{255}{\sqrt{MSR(A,B)}}\right) \qquad (9)$$

All these similarity measure and the classical measures for images comparison (*MSR*, *PSNR*) will be compared with the proposed similarity measure using real data.

3. The proposed similarity measures

Definition 3.1. For A,B ∈ F(X), we define

$$S^r(A,B) = \frac{2^r a.b}{a.a + b.b + (2^r - 2)a.b} = \frac{2^r \sum_{i=1}^{n} a_i.b_i}{\sum_{i=1}^{n} a_i^2 + \sum_{i=1}^{n} b_i^2 + (2^r - 2)\sum_{i=1}^{n} a_i.b_i}, \quad r \ge 0 \qquad (10)$$

Definition 3.2. . For A,B ∈ F(X), we define

$$\overline{S}^r(A,B) = \frac{2^{-r} a.b}{a.a + b.b + (2^{-r} - 2)a.b} = \frac{2^{-r} \sum_{i=1}^{n} a_i.b_i}{\sum_{i=1}^{n} a_i^2 + \sum_{i=1}^{n} b_i^2 + (2^{-r} - 2)\sum_{i=1}^{n} a_i.b_i}, \quad r \ge 0 \qquad (11)$$

We call $S^r(A,B)$ and $\overline{S}^r(A,B)$ similarity measures between fuzzy sets A and B. These two classes of similarity measures can be considered as a general form of the similarity measure introduced in [2] by putting *r=0*. Their properties will be investigated under the definition 2.1. The following three lemmas will be needed to do this.

Lemma 3.1. Let $p, q \in [0,1]$ and $p \geq q$. Then $\frac{p}{1+p} \geq \frac{q}{1+q}$

Lemma 3.2. Let $0 \leq p \leq 1$. Then $0 \leq \frac{2p}{1+p} \leq 1$.

Lemma 3.3. Let $p, q \in [0,1]$ and $p \geq q$. Then $\frac{p}{2-p} \geq \frac{q}{2-q}$

In the following the properties of $S^r(A,B)$ under the definition 2.1 will be investigated and the same thing can be done for $\overline{S}^r(A,B)$ with using lemma 3.3.

SP1. A,B ∈ F(X), one have $S^r(A,B) = S^r(B,A)$
The proof is obvious.

SP2. $S^r(D,D^c) = 0$, if D is a scrip set
The proof is obvious.

SP3. $S^r(E,E) = \max_{A,B \in F(X)} S^r(A,B)$, for all $E \in F(X)$
The proof is obvious.

SP4. If $A \subseteq B \subseteq C$ for all $A,B,C \in F(X)$
then $S^r(A,B) \geq S^r(A,C)$ and $S^r(B,C) \geq S^r(A,C)$.

Proof. Mathematical induction will be used to check the proof of this proposition

(i) For *r = 0*

$$S^0(A,B) = \frac{\sum_{i=1}^{n} a_i.b_i}{\sum_{i=1}^{n} a_i^2 + \sum_{i=1}^{n} b_i^2 - \sum_{i=1}^{n} a_i.b_i}$$

which is the same $S_k(A,B)$ introduced in [2] and the proof of this case is given in details in [2].

(ii) For $r = k$, we suppose the relation is correct i.e.
If $A \subseteq B \subseteq C$ for all $A,B,C \in F(X)$
then $S^k(A,B) \geq S^k(A,C)$ and $S^k(B,C) \geq S^k(A,C)$.

(iii) We will try to prove that the relation is true at $r = k+1$ using case (ii)
i.e. If $A \subseteq B \subseteq C$ for all $A,B,C \in F(X)$
Then $S^{k+1}(A,B) \geq S^{k+1}(A,C)$ and $S^{k+1}(B,C) \geq S^{k+1}(A,C)$.

$$S^{k+1}(A,B) = \frac{2^{k+1}\sum_{i=1}^{n} a_i.b_i}{\sum_{i=1}^{n} a_i^2 + \sum_{i=1}^{n} b_i^2 + (2^{k+1}-2)\sum_{i=1}^{n} a_i.b_i}$$

$$= \frac{2\left(\frac{2^k\sum_{i=1}^{n} a_i.b_i}{\sum_{i=1}^{n} a_i^2 + \sum_{i=1}^{n} b_i^2 + (2^k-2)\sum_{i=1}^{n} a_i.b_i}\right)}{1+\left(\frac{2^k\sum_{i=1}^{n} a_i.b_i}{\sum_{i=1}^{n} a_i^2 + \sum_{i=1}^{n} b_i^2 + (2^k-2)\sum_{i=1}^{n} a_i.b_i}\right)} = \frac{2\,S^k(A,B)}{1+S^k(A,B)}$$

Using lemma 3.1 and case (ii) $S^k(A,B) \geq S^k(A,C)$, we have

$$S^{k+1}(A,B) = \frac{2\,S^k(A,B)}{1+S^k(A,B)} \geq \frac{2\,S^k(A,C)}{1+S^k(A,C)} = S^{k+1}(A,C)$$

i.e. $S^{k+1}(A,B) \geq S^{k+1}(A,C)$ for all $A,B,C \in F(X)$

In the same way one can easily prove that
$S^{k+1}(B,C) \geq S^{k+1}(A,C)$ for all $A,B,C \in F(X)$.
Therefore, If $A \subseteq B \subseteq C$ for all $A,B,C \in F(X)$
then $S^r(A,B) \geq S^r(A,C)$ and $S^r(B,C) \geq S^r(A,C)$.

There are some other similarity properties which should be investigate such as:

SP5. For $A,B \in F(X)$ one should have $S^r(A,B) = 1$ if A=B.
The proof is obvious

SP6. For $A,B \in F(X)$ one should have $0 \leq S^r(A,B) \leq 1$.

Proof. The proof of this case can be done using mathematical induction as follows:

(a) For *r* = 0, we have $S^0(A,B) = S_k(A,B)$ and the proof of this case founded in [2].
i.e. $0 \leq S^0(A,B) = S_k(A,B) \leq 1$, for $A,B \in F(X)$.

(b) For $r = k$, we suppose the relation is correct
i.e. $0 \leq S^k(A,B) \leq 1$, for $A,B \in F(X)$.

(c) We will try to prove that the relation is true at $r = k+1$ using case (b)
i.e. $0 \leq S^{k+1}(A,B) \leq 1$, for $A,B \in F(X)$.
From proof of **SP4** we have

$$S^{k+1}(A,B) = \frac{2\,S^k(A,B)}{1+S^k(A,B)}$$

Using lemma 2.2 and case (b)
$0 \leq S^k(A,B) \leq 1$, for $A,B \in F(X)$, we have

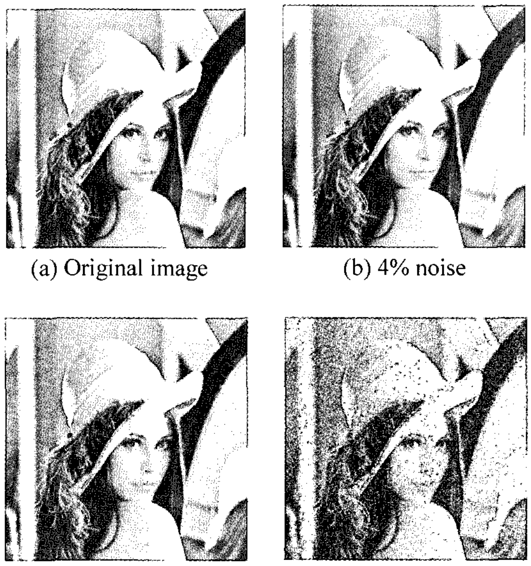
$$0 \leq S^{k+1}(A,B) = \frac{2\,S^k(A,B)}{1+S^k(A,B)} \leq 1$$

Therefore, $0 \leq S^r(A,B) \leq 1$, for $A,B \in F(X)$.

4. Experimental results

The performance of the new similarity measures in comparison with the mentioned similarity measures and with the classical measures for images comparison, the mean square error (*MSE*), and the peak-signal-to-noise-ratio (*PSNR*) will be illustrated in this section. In order to do this, two experiments are carried out.

In the first experiment three different percentages (4%,10%, 33%) of gaussian noise are added to the original image used in the test. The original image (Lena) and the noisy images with different percentages are shown in Fig.1. The results of this experiment are shown in Table 1. This experiment show how the proposed similarity measures react to gaussian noise in comparison with other similarity measures. From the results of table (1) one can note that the proposed similarity measures don't affect too much due to noise because the noisy image is coming from the original image and it has to be similar to the original one. Also, the value of proposed similarity measures decreases with respect to an increasing noise percentage.



(a) Original image (b) 4% noise

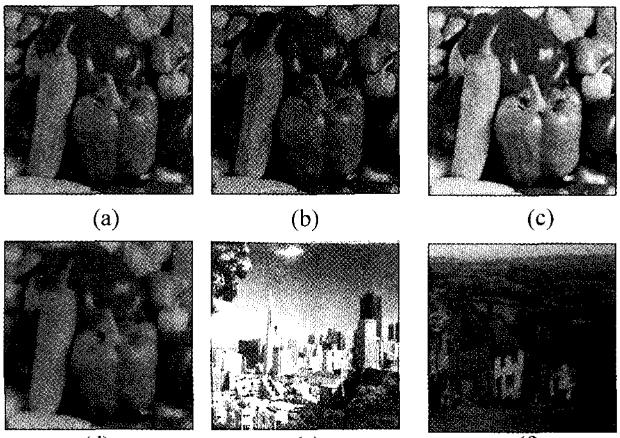
(c)10% noise (d) 33% noise

Fig. 1. Lena image with different percentages(4%,10%, 33%) of gaussian noise.

Table 1. Performance of measures applied to two images (A is original, and B is with three different percentages (4%,10%, 33%) of gaussian noise).

	4%	10%	33%
S_1 (A,B)	0.98329	0.96063	0.83420
S_2 (A,B)	0.98622	0.96724	0.81340
S_3 (A,B)	0.94510	0.87451	0.60784
S_4 (A,B)	0.97285	0.93659	0.82379
S_5 (A,B)	0.96697	0.92437	0.79686
S_6 (A,B)	0.98624	0.96766	0.81398
S_7 (A,B)	0.99902	0.99524	0.90359
S^0 (A,B)	0.99900	0.99102	0.83214
S^r (A,B)	0.99950	0.99334	0.84032
$\overline{S}^r$ (A,B)	0.99801	0.98822	0.85947
MSE(A,B)	10.16	100.77	871.64
PSNR(A,B)	35.54	28.12	18.73

The second experiment has been performed to test the sensitivity of the proposed measures in comparison with the other mentioned measures using images with different types of distortions. In this case 'peppers' image is used, a variety of corruption is add: JPEG-compression, enlightening, blur, and other different types of images (San Francisco, hill). The original and distorted images are shown in Fig. 2. In Table 2 the value of all measures are calculated with respect to the prototype image (a) which is the leftmost one in Fig. 2. Based on the S_6 and (S_1, S_7), one can make an incorrect conclusion that images (a) and (e), (a) and (f) (Fig.2) have the same similarity respectively, while the proposed similarity measures show larger difference between these images. According to the results of table 2, it is easy to note that the performance of MSE is extremely poor in the sense that images are nearly identical. On the other hand the proposed measures show better sensitivity in comparison of similar images and better separation between different images, with $\overline{S}^r$ is more sensitive to noise and dissimilarity than S^r.



(a) (b) (c)

(d) (e) (f)

Fig. 2. (a) The original 'pepers' image, (b) JPEG-compression, (c) enlightened, (d) blurred, (e) san. image, (f) hill image.

Table 2. Performance of measures applied to 'peppers' image with a variety of corruption and other different images.

	(a) vs. (b)	(a) vs. (c)	(a) vs. (d)	(a) vs. (e)	(a) vs. (f)
S_1 (A,B)	0.93120	0.99023	0.97113	0.72359	0.91214
S_2 (A,B)	0.94981	0.96502	0.96789	0.66685	0.76309
S_3 (A,B)	0.81176	0.89974	0.90112	0.21176	0.63529
S_4 (A,B)	0.91221	0.90565	0.96634	0.50011	0.61685
S_5 (A,B)	0.89930	0.93881	0.97126	0.55550	0.62628
S_6 (A,B)	0.92321	0.98302	0.98330	0.93738	0.78754
S_7 (A,B)	0.96034	0.97661	0.97354	0.46604	0.94007
S^0 (A,B)	0.92232	0.97893	0.96780	0.56245	0.61203
S^r (A,B)	0.93372	0.96972	0.97560	0.55332	0.73215
$\overline{S}^r$ (A,B)	0.90865	0.93991	0.95214	0.20012	0.38924
MSE(A,B)	128.22	103.25	120.23	2968.2	1062.6

Finally, only a qualitative evaluation of the proposed similarity measures performance is made in the present work, because there are no formal criteria to compare various similarity measures and to define a universal measure (application independent). The only disadvantage of the proposed similarity measures is that like all the other measures the proposed similarity measures have the localization of their usage, i.e. they may provide a useful way to measure the similarity in some cases. Hence it is very hard to say that the proposed measures are the best in all cases.

5. Conclusions

Measuring the similarity between fuzzy sets plays a vital role in several fields. Constantly looking for a better similarity measure method is a pursuing of fuzzy mathematicians, however, none of all well-known similarity measure methods is all-powerful, and all have the localization of its usage. In the present paper we proposed a new class of similarity measures and examined its properties. The proposed class of similarity measure was compared with other similarity measures using real data. Because of it have some good properties, it can be concluded that the proposed generalized similarity measure could improve distinguish precision and enhance the capability of classification of some similar sets, therefore, it can provide a useful way to measure the similarity between fuzzy sets.

6. References

[1] H. Muller, N. Michoux, D. Bandon, A. Geissbuhler, A review of content-based image retrieval systems in medical applications-clinical benefits and future directions, *Medical Informatics*, vol. 73, pp. 1-23, 2004.

[2] 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.

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

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

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

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

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

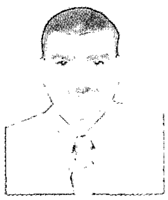
[8] D. Van der Weken, M. Nachtegael , E.E. Kerre, An overview of similarity measures for images, Proceeding of ICASSP 2002 (IEEE International Conference on Acoustics, Speech and Signal processing), USA, pp.3317-3320, 2002.

[9] 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.

[10]W. Wang, X. Xin, Distance measure between intuitionistic fuzzy sets, *Pattern Recognition Letters*, vol. 26, pp. 2063-2069, 2005.

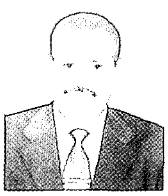
[11]C. Zhang, H. Fu, Similarity measures on three kinds of fuzzy sets, *Pattern Recognition Letters*, vol. 26, 2006, (article in press).

[12]L. A. Zadeh, Fuzzy sets, *Inform. Control*, vol. 8, pp. 338-353, 1965.



Salah Omran received his Ph.D. degree in Mathematics in 2005 from University of Muenster, Germany. He is a lecturer in the Department of Mathematics at South Valley University, Egypt. His research interests are operator algebras, k-theory, and fuzzy sets and its applications.

e-mail : salahomran@yahoo.com
Tel : +2 0102594601.
Fax : +2 0965211279.



M. Hassaballah received his B.Sc. degree in Mathematics in 1997 and M.Sc. degree in Computer Science in 2003, all from South Valley University, Egypt. He is an assistant lecturer in the Department of Mathematics at South Valley University, Egypt. His research interests are in computer vision, fractal image compression, content-based image retrieval, similarity measures, and high performance computing.

e-mail : mahcs2010@yahoo.ca
Tel : +2 0965224219.
Fax : +2 0965211279.