Some Notes on Fuzzy Similarity Measures and Application to Classification of Shapes, Recognition of Arabic Sentences and Mosaic

Leila Baccour, Adel M. Alimi, and Robert I. John

***Abstract*—Fuzzy Similarity measures are used to compare different kinds of objects such as images. Their definitions are based on proximity measures, operations on fuzzy sets etc. which makes different propositions of properties of similarity measures. Consequently, the latter are not common for all similarity measures presented in literature. In this paper we present and discuss the properties of fuzzy similarity and distance measures. We present fuzzy similarity measures from literature and we discuss their validation to the common existing properties.**

In order to compare and show the differences between fuzzy similarity measures, we apply them to three applications using different data sets defined with fuzzy attributes. The obtained results are good for most existing similarity measures, but some measures give best results and some others worst results. Therefore, relationships between obtained results and measures validation to their properties are discussed to know the influence of some properties on the results.

***Index Terms*—similarity measures, distance measures, fuzzy sets, Arabic sentence recognition, classification of shapes, similarity measure properties, distance measure properties, mosaic recognition.**

I. Introduction

In literature, while some similarity measures are proposed, properties of fuzzy similarity and fuzzy distance measures are still not common for all proposed measures even though studied and discussed. The studies on fuzzy similarity measures are mostly theoretic [1]–[5] and were the subject of several publications. [6]–[8] applied fuzzy similarity measures to image processing. [9] presented a fuzzy similarity inference method for fuzzy reasoning. [10] proposed and applied a similarity measure to shape retrieval using the SQUID data set described with Fourier descriptor. [11] aggregated implication operators to a similarity measure applied to shape classification and [12], [13] applied fuzzy similarity measures to handwritten Arabic sentences recognition. [14] presented a recognition system for handwritten Latin words described with Gabor filters and used a process recognition based on fuzzy logic to study the variations of handwriting style of different writers. [15] presented an off-line Arabic signature recognition and verification using fuzzy concepts, other works are found in [16]–[19]. In these works, similarity measures are applied or proposed, some comparisons between some measures are done in some specific domains like image processing [6], [7] but still specific to some treatments on images.

Similarity measures between fuzzy sets are numerous. It is not trivial to predict the measure that provides the best results in any application. Therefore, numeral comparisons between fuzzy similarity measures (FSMs) are important to show experimental differences between them.

In this paper, we sum up common existing properties of Fuzzy similarity and distance measures. We apply FSMs, from literature, to classification of shapes, mosaic recognition and Arabic sentence recognition. These applications allow us to find the measures giving best results for used data sets. Consequently, selection of similarity measures can be easy for application in other topics. Experimental results are discussed and deductions are made according to the measures results and their validation to the properties.

This paper is organized as follows: In the second section we present preliminaries, followed by presentation, discussion and summary of common existing properties for similarity and distance measures. In the forth section we expose FSMs from literature, followed by their validation to common properties. In the sixth , seventh and eighth sections we apply fuzzy similarities respectively to classification of shapes, mosaic recognition and Arabic sentence recognition. In last sections, we compare results and we give conclusions.

II. Preliminaries

Let X the discourse universe, a sample $x \in X$ having p attributes as: $x = (x_1, x_2, \ldots, x_p)$

- Let $P(X)$ the set of all crisp sets in X, $FS(X)$ the set of all fuzzy sets in X, A and B tow fuzzy sets in X defined as: $A = \{(x, \mu_A(x)) | x \in X, \mu_A(x) \in [0,1]\}$, $B = \{(x, \mu_B(x)) | x \in X, \mu_B(x) \in [0,1]\}$
- Let $A^c \in FS(X)$ the complement of $A \in FS(X)$ and $\mu_{A^c}(x) = 1 - \mu_A(x), \forall x \in X$.
- The intersection and the union operators used for all formulas are the minimum and the maximum proposed in [20].
- The cardinality of a fuzzy set A in a universe X is defined as: $|A| = \sum_{x \in X} \mu_A(x)$

III. Distance Measures versus Similarity Measures

A measure of similarity or dissimilarity defines the resemblance between two samples or objects. For a similarity measure the resemblance is more important when its value increases. However, for a dissimilarity measure (i.e.

Supported by the grants from the General Direction of Scientific Research (DGRSRT), Tunisia, under the ARUB program.

L. Baccour and A. M. Alimi are with REGIM-Lab.: REsearch Groups in Intelligent Machines, University of Sfax, ENIS, BP 1173, Sfax, 3038, Tunisia. (leilabaccour@gmail.com; adel.alimi@ieee.org)

R. I. John is with Automated Scheduling, Optimisation and Planning (ASAP) group, Room C72 Computer Science Jubilee Campus Wollaton Road Nottingham, NG8 1BB, UK (Robert.John@nottingham.ac.uk)

(Advance online publication: 27 May 2014)

a distance measure) the resemblance is stronger when its value decreases. So, similarity and distance measures are dual concepts, if they are normalized: similarity measure= 1-distance measure and vice versa.
The properties of FSMs are discussed in some works as [4], [21], [22] which presented properties based on proximity measures. In addition [9] and [23] presented some modification on these properties and provided some other properties. In the following, we discuss the properties of fuzzy similarity and distance measures. Our discussion begins with the presentation of crisp distance measure properties.

A. Crisp Distance Measure Properties

A distance measure is a metric d defined as a function: $d : X^2 \to \mathbb{R}^+$ and have the following properties:

1) $\forall\ x \in X, d(x, x) = 0$
2) $\forall\ x, y \in X, d(x, y) = 0 \Rightarrow x = y$
3) $\forall\ x, y \in X, d(x, y) = d(y, x)$
4) $\forall\ x, y, z \in X, d(x, y) \leq d(x, z) + d(z, y)$

B. Fuzzy Distance Measure Properties

In [2] the following properties of a fuzzy distance measure between two sets A and B are defined.

1) $d(A, B) = d(B, A)\ \forall\ A, B \in FS(X)$
2) $d(A, A) = 0, \forall A \in FS(X)$
3) $d(D, D^c) = max_{A,B \in FS(X)} d(A, B)$ for all $D \in P(X)$
4) $\forall\ A, B, C \in FS(X),$ if $A \subseteq B \subseteq C$ then $d(A, B) \leq d(A, C)$ and $d(B, C) \leq d(A, C)$

C. Fuzzy Similarity Measure Properties

1) Definition 1: The properties of a fuzzy similarity measure $S(A, B)$ between two sets A and B of $FS(X)$ are proposed in [4], [21], [22], based on properties of proximity measures, as follows:

1) $S(A, B) = S(B, A)\ \forall\ A, B \in FS(X)$
2) $S(E, E) = max_{A,B \in FS(X)} S(A, B)$ for all $E \in FS(X)$
3) $S(D, D^c) = 0$, if D is a crisp set
4) $\forall\ A, B, C \in FS(X),$ if $A \subseteq B \subseteq C$ then $S(A, B) \geq S(A, C)$ and $S(B, C) \geq S(A, C)$

[21] noted that if a similarity measure satisfies $max_{A,B \in FS(X)} S(A, B) = 1$ then it is called normal similarity measure on $FS(X)$. Similarly, if a distance measure satisfies $max_{A,B \in FS(X)} d(A, B) = 1$ then it is called normal distance measure on $FS(X)$.

2) Definition 2: In [9] the following definition of fuzzy similarity measure properties is given:

1) $S(A, B) = S(B, A)$
2) $A = B \Rightarrow S(A, B) = X$
3) $S(X, \emptyset) = \emptyset$
4) $\forall\ A, B \in FS(X)$, if $A \subset B \subset C$ then $S(A, C) \leq min(S(A, B), S(B, C))$.

The author consider that the similarity between two objects can not be limited to a number between 0 and 1 because it can not represent well the knowledge, for example, a similarity between two persons can be: "a little" or "very resemble". This is expressed by the second property where the similarity between two fuzzy sets is the whole discourse universe. We find that this property is ambiguous and represent limitation for the use of similarity measures in recognition or classification systems.

3) Definition 3: In [23] the following definition of fuzzy similarity measure properties is given:

1) $S(B, A) = S(A, B)$
2) $0 \leq S(A, B) \leq 1$
3) $S(A, B) = 1$ if and only if $A = B$
4) $\forall\ A, B \in FS(X)$, if $A \subset B \subset C$ then $S(A, C) \leq min(S(A, B), S(B, C))$.
5) For two fuzzy sets A and B not null, if $S(A, B) = 0$ then $min(\mu_A(x), \mu_B(x)) = 0\ \forall\ x \in FS(X)$

D. Notes on Fuzzy Similarity and Distance Properties

The propositions of similarity measure properties are varied. Measures in literature do not validate all given definitions. In this section we summarize all common presented properties for similarity and distance measures. Therefore, we present definitions of similarity and distance measures with most common properties. In the following we discuss all properties definitions for similarity and distance measures.

- The forth property exists in all given definitions of similarity measures. It corresponds to the fourth property in distance measures. Thus, this is a common property in all definitions.
- For the third property proposed in [21] (definition 1) $max_{A,B \in FS(X)} S(A, B) = 1$ for normal similarity measures, it corresponds the third property cited in definition 3. Also, This property much properties for crisp distance measures. However, it is not present in all definitions of FSMs and is replaced in definition 1 with the comparison between crisp sets and their complements as:
  $S(D, D^c) = 0$, if D is a crisp set
  We find that this comparison is not interesting for comparing two fuzzy sets (i.e, we did not find the relation between the comparison of two fuzzy sets with comparing crisp sets to their complements). Therefore, the most suitable third property is given in [21], [23] (definition 1, definition 3):
  (3) $\forall\ A, B \in FS(X), S(A, B) = 1 \Longleftrightarrow A = B$
  The third property in distance measures definition is given as:
  $d(D, D^c) = max_{A,B \in FS(X)} d(A, B)$ for all $D \in P(X)$
  This property compare between crisp sets and their complements. For the same reason given previously and according to the duality between fuzzy distance and similarity measures, the third property of distance measures will be:
  (3) $\forall\ A, B \in FS(X), d(A, B) = 0 \Longleftrightarrow A = B$
  We note that the third property proposed in [9], definition 2, is obvious.
- For the second property of fuzzy distance measures [2] (see Sect. III-B)
  $d(A, A) = 0, \forall A \in FS(X)$ it corresponds the second property of similarity measures proposed in [4], definition 1: $S(E, E) = max_{A,B \in FS(X)} S(A, B), \forall A \in$

$FS(X)$ [22]. This property is not common for all definitions of FSMs.
If we rely on duality between similarity and distance measures, we conclude that their properties must be complementary and not contradictory. In addition, properties presented in definition 1 are based on proximity measures and can not be very suitable for FSMs. Thus, according to the cited reasons, the second property of FSMs will be:
(2) $S(A, A) = 1, \forall A \in FS(X)$
However this property is redundant for both similarity and distance measures and can be deduced from the third property. In addition, the similarity and distance measures must be positive and if they are normalized they must be less or equal 1, this is mentioned in definition 3. Thus, the second property of FSMs will be:
(2) $0 \leq S(A, B) \leq 1, \forall A \in FS(X)$
and the second property of distance measures will be:
(2) $0 \leq d(A, B) \leq 1, \forall A \in FS(X)$

- The first property is present in all proposed definitions of FSMs, and it corresponds to the first property of distance measures.
- The fifth property proposed in definition 3 indicates that intersection must be 0 if the sets are dissimilar. This property is included in the common third property.

To sum up, the common existing properties of fuzzy similarity measures are:

1) $S(A, B) = S(B, A)\ \forall\ A, B \in FS(X)$
2) $0 \leq S(A, B) \leq 1, \forall A \in FS(X)$
3) $\forall\ A, B \in FS(X), S(A, B) = 1 \iff A = B$
4) $\forall\ A, B, C \in FS(X),$ if $A \subseteq B \subseteq C$ then $S(A, B) \geq S(A, C)$ and $S(B, C) \geq S(A, C)$

The properties of fuzzy distance measures are:

1) $d(A, B) = d(B, A)\ \forall\ A, B \in FS(X)$
2) $0 \leq d(A, B) \leq 1, \forall A \in FS(X)$
3) $\forall\ A, B \in FS(X), d(A, B) = 0 \iff A = B$
4) $\forall\ A, B, C \in FS(X),$ if $A \subseteq B \subseteq C$ then $d(A, B) \leq d(A, C)$ and $d(B, C) \leq d(A, C)$

We note that we did not give new properties in these definitions. However, we retain all common properties in existing definitions and we altered some of them according to the duality concept between similarity and distance measures.

IV. Similarity Measures between Fuzzy Sets From Literature

In the following, we present FSMs defined in literature. We mention that in all formulas μ_A^i and μ_B^i are respectively the membership degree of the i^{th} elements in sets A and B of n elements.

A. Measures Based on Operations on Fuzzy Sets

In the sequel, similarity measures based on union, intersection and cardinality operations on fuzzy sets are presented. [4] proposed this measure:

$$M(A, B) = \frac{\sum_{i=1}^{n} min\left(\mu_A^i, \mu_B^i\right)}{\sum_{i=1}^{n} max\left(\mu_A^i, \mu_B^i\right)} \tag{1}$$

[1] proposed the following similarity measures between fuzzy values:

$$T(A, B) = max_i(min(\mu_A^i, \mu_B^i)) \tag{2}$$

$$P = \frac{\sum_{i=1}^{n} \mu_A^i \cdot \mu_B^i}{max\left(\sum_{i=1}^{n} \mu_A^i \cdot \mu_A^i, \sum_{i=1}^{n} \mu_B^i \cdot \mu_B^i\right)} \tag{3}$$

[22] proposed the following similarity measure:

$$S_1(A, B) = \frac{1}{n} \sum_{i=1}^{n} \left(\frac{min\left(\mu_A^i, \mu_B^i\right)}{max\left(\mu_A^i, \mu_B^i\right)} \right) \tag{4}$$

If the denominator equalizes to zero then $S_1(A, B) = 1$.
In [8] these similarity measures are presented.

$$S_4(A, B) = \frac{|A^c \cap B^c|}{|A^c \cup B^c|} \tag{5}$$

$$S_5(A, B) = \frac{|A \cap B|}{max(|A|, |B|)} \tag{6}$$

$$S_6(A, B) = \frac{|A^c \cap B^c|}{max(|A^c|, |B^c|)} \tag{7}$$

$$S_7(A, B) = \frac{|A \cap B|}{min(|A|, |B|)} \tag{8}$$

$$S_8(A, B) = \frac{|A^c \cap B^c|}{min(|A^c|, |B^c|)} \tag{9}$$

In [9], this similarity measure is proposed.

$$S_3(A, B) = \sum_{i=1}^{n} \frac{2 * min(\mu_A^i, \mu_B^i)}{\mu_A^i + \mu_B^i} \tag{10}$$

if $\mu_A^i + \mu_B^i = 0$ then $S_3(A, B) = 1$.
To be normalized, this measure must be divided by n.

B. Measures Based on Symmetric Difference of Fuzzy Sets

In [8] similarity measures based on symmetric difference of two fuzzy sets are presented. First, the authors defined the difference between two fuzzy sets A and B as: $A \setminus B = A \cap B^c$. Thus, they defined the symmetric difference as: $A\Delta B = (A \setminus B) \cup (B \setminus A)$. The proposed measures are:

$$S_9(A, B) = \frac{|(A\Delta B)^c|}{max(|(A \setminus B)^c|, |(B \setminus A)^c|)} \tag{11}$$

$$S_{10}(A, B) = \frac{|(A\Delta B)^c|}{min(|(A \setminus B)^c|, |(B \setminus A)^c|)} \tag{12}$$

$$S_{11}(A, B) = 1 - \frac{1}{n} \sum_{i=1}^{n} (A\Delta B)(x_i) \tag{13}$$

C. Measures Based on Distance Measures

[4] proposed The following similarity measures:

$$L(A, B) = 1 - max_i\left(|\mu_A^i - \mu_B^i|\right) \tag{14}$$

$$S(A, B) = 1 - \frac{\sum_{i=1}^{n} |\mu_A^i - \mu_B^i|}{\sum_{i=1}^{n} (\mu_A^i + \mu_B^i)} \tag{15}$$

[1] studied the measures proposed by [4] and demonstrated that the measure (14) is based on the geometric distance model proposed in [24] as:

$$d_r(A, B) = \left(\sum_{i=1}^{n} (a_i - b_i)^r \right)^{\frac{1}{r}} \tag{16}$$

where $a = (a_1, a_2, \ldots, a_n)$ and $b = (b_1, b_2, \ldots, b_n)$ are two points in a n-dimensional space. When r approaches ∞ d_r becomes:

$$d_\infty = max_i |a_i - b_i| \tag{17}$$

Then, [1] proposed the following measure between fuzzy values:

$$W(A,B) = 1 - \frac{\sum_{i=1}^{n} |\mu_A^i - \mu_B^i|}{n} \tag{18}$$

[22] proposed the following similarity measure:

$$S_2(A,B) = \frac{1}{n} \sum_{i=1}^{n} \left(1 - |\mu_A^i - \mu_B^i|\right) \tag{19}$$

[25] presented the following distance measure which is a divergence measure between two fuzzy sets A and B:

$$\begin{aligned} D_E(A,B) = \sum_{i=1}^{n} \Big(2 - \left(1 - \mu_A^i + \mu_B^i\right) e^{(\mu_A^i - \mu_B^i)} \\ - \left(1 - \mu_B^i + \mu_A^i\right) e^{(\mu_B^i - \mu_A^i)}\Big) \end{aligned} \tag{20}$$

where $0 \leq D_E(A,B) \leq n(2 - 2e^{-1})$
and demonstrated that the following divergence measure presented in [26] is a distance measure.

$$\begin{aligned} D_L(A,B) = \sum_{i=1}^{n} \Big(\left(\mu_A^i - \mu_B^i\right) \ln \frac{1 + \mu_A^i}{1 + \mu_B^i} \\ + \left(\mu_B^i - \mu_A^i\right) \ln \frac{2 - \mu_A^i}{2 - \mu_B^i} \Big) \end{aligned} \tag{21}$$

Therefore, [25] presented the following distance measure:

$$\begin{aligned} d(A,B) = 1 - \frac{1}{n} \sum_{i=1}^{n} \\ \frac{max(min(\mu_A^i, \mu_B^i), min(\mu_{A^c}^i, \mu_{B^c}^i))}{max(max(\mu_A^i, \mu_B^i), max(\mu_{A^c}^i, \mu_{B^c}^i))} \end{aligned} \tag{22}$$

In [8] a similarity measure using the distance (21) is presented as:

$$\begin{aligned} S_{D_L(A,B)} = 1 - \frac{1}{2\ln 2} \sum_{i=1}^{n} \Big(\left(\mu_A^i - \mu_B^i\right) \ln \frac{1 + \mu_A^i}{1 + \mu_B^i} \\ + \left(\mu_B^i - \mu_A^i\right) \ln \frac{2 - \mu_A^i}{2 - \mu_B^i} \Big) \end{aligned} \tag{23}$$

D. Measures Based on Implication Operators

Similarity measures based on implication operators are proposed in [5] as follows:

$$E_p = min(inf_{i \in \{1 \ldots n\}} p(a_i, b_i), inf_{i \in \{1 \ldots n\}} p(b_i, a_i)) \tag{24}$$

Implication operators which can be used in (24) are:

- Lukasiewiez implication operator:
  $Min(1 - a + b, 1)$
- Gödel implication operator:
  $\begin{cases} 1 & \text{if} \quad a \leq b \\ b & \text{else} \end{cases}$
- Goguen implication operator:
  $\begin{cases} min(\frac{b}{a}, 1) & \text{if} \quad a \neq 0 \\ 1 & \text{else} \end{cases}$

Similarity measures obtained with each implication are called respectively S_L, S_{GO} and S_G.

E. Notes on Presented Measures

The presented measures from literature are based on some operators on fuzzy sets. Some of them are based on distance measures between crisp sets. We cite the measures W (18) and S_2 (19) which are dual concepts of Hamming distance ($W = 1-$Hamming distance). We recall that Hamming distance between two crisp vectors x and y is defined as: $d(x,y) = \frac{1}{n} \sum_{i=1}^{n} |x_i - y_i|$.
According to its frequency of use in most systems treating crisp values, we will use the Euclidean distance between two fuzzy sets, based on the distance d_r (16) with $r = 2$, defined as:

$$d_2(A,B) = \frac{1}{\sqrt{n}} \sqrt{\sum_{i=1}^{n} \left(\mu_A^i - \mu_B^i\right)^2} \tag{25}$$

V. Measures Properties Validity

In this section, we verify the satisfaction of the measures presented above to the properties of similarity and distance measures deduced and presented in Sect. III-D.
It is obvious that all measures verify the properties 1 and 2. Thus, we present the properties 3 and 4 verification for each measure. If the measure is a distance measure the properties of distance are verified, inversely, if the measure is a similarity, the properties of similarity are verified. In table I below, we present the results of validation of the measures to the properties noted P3 and P4. We note that we indicate "yes" if the property is verified and we indicate "no" otherwise.

TABLE I
Validation of Fuzzy Similarity Measures to Common Properties

FSMs	P3	P4
M(1), L (14), S_2 (19), S_G (24) W(18), S_4 (5), S_L (24) S_5 (6), S_6 (7), d_2 (25)	yes	yes
S (15), P (3), S_{D_L} (23) d (22), S_3 (10) S_1 (4), D_E (20), S_9 (11)	yes	no
T (2), S_{11} (13) S_7 (8), S_8 (9) S_{10}(12), S_{GO} (24)	no	no

VI. Classification of Shapes

A. Presentation of Shapes Data Set

We use the SQUID data set, which contains about 1100 images of marine creatures (fish). Each image shows a distinct species on a uniform background and is processed to recover the boundary edge, which is then represented by extended Curvature Scale Space (CSS) descriptors.
The CSS descriptors introduced by [27], permit to register the concavities of a curve through successive filtering. They have the advantage of being invariant to scale, translation, and rotation, and are shown to be robust and tolerant of noise, however, they are inadequate to represent the convex segments of a shape. To solve this problem, [28] presented a solution to remedy the inability of the CSS descriptors to represent convex segments of a shape. The proposed idea is to create a dual shape of the input shape where all convex segments are transformed to concave segments.
The CSS descriptors are considered as local features and do

not capture the global features of an image edge. Thus, two extra global features: circularity and eccentricity are used. for more details see [11], [29].
In conclusion, the features used to describe the shapes data set are called "extended CSS descriptors" and are:

- Circularity
- Eccentricity
- CSS descriptors of original shape
- CSS descriptors of dual shape

All shapes are labeled with the clusters in which they are assigned. We note thirty classes of shapes.

B. Constitution of Fuzzy Features of Shapes

In this section, we are interested in creating fuzzy features of shapes. This stage permits us to pass from a crisp data set constructed by crisp values of extended CSS descriptors to a fuzzy data set. The last-mentioned data contains values between 0 and 1, appointing membership degrees to fuzzy sets. To choose membership functions, we divided the data set of shapes into two data sets: the first which is constituted of 738 shapes serves as reference data set and the second which is constituted of 362 shapes serves as test data set. Every shape, in the test data set, corresponds to the shapes of same cluster in the reference data set (i.e., the number of reference shapes of the same cluster of the test shape is between one and fifteen). The process of fuzzification is described in figure 1 [12]. We specify that The features of every shape are: one value of eccentricity, one value of circularity, a vector of concavity points (abscissas, ordinates) and a vector of convexity points. We note, that one shape can take different positions, so, two shapes can be similar if concavity ordinates are equal independently of abscissas (i.e. an abscissa can be on the right and the other on the left) see figure 2. Consequently, abscissas are not important to retain for features concavity and convexity. Therefore, we fuzzify for each shape : eccentricity, circularity, ordinates of concavity points and ordinates of convexity points.

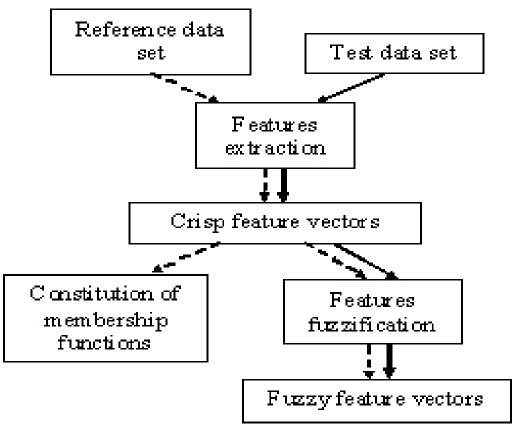
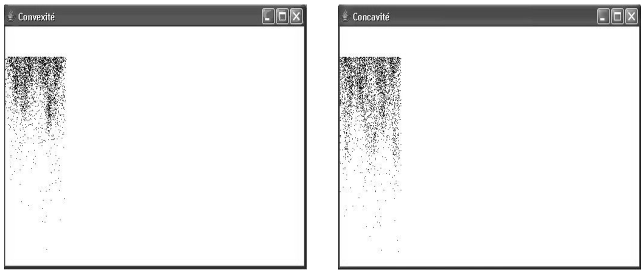


Fig. 1. Process of Features Fuzzification

Fig. 2. Different Positions of a Shape

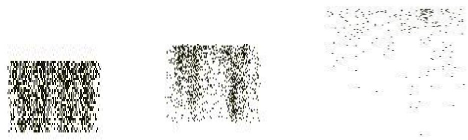
The reference data set is used to constitute membership functions for each feature (figure 1). Therefore, we present each feature in the plan to determine its linguistic variables (i.e. linguistic variables corresponding to each membership function). This is shown in figure 3 and figure 5 [12].



(a) (b)

Fig. 3. Representation in the Plan of (a) Ordinates of Convexity Points and (b) Ordinates of Concavity Points of Data Set of Reference Shapes

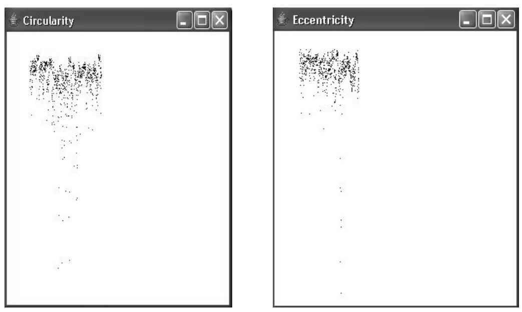
Figure 3-(a) shows that the convexity points are less frequent than the concavity points on figure 3-(b). The first figure has two peaks and the second has three peaks. However, the condensation of the points is the same. So, if we examine (a) and (b) in figure 3, we can conclude that we can divide each of them, in the top-down direction, into three sets representing low, medium and high convexity or concavity.
Figure 4 [12], represents the sets obtained after the division of convexity sets and shows the low convexity figure 4-(a), the medium convexity figure 4-(a) and the high convexity, figure 4-(c). We note that the points ordinates of convexity and concavity are represented from the lower value to the higher value in the top-down sense of the figure.



(a) (b) (c)

Fig. 4. (a) Low convexity (b) Medium convexity (c) High Convexity

Figure 5 represents the features circularity and eccentricity. It is easy to remark that the eccentricity can be divided, in the top-down sense, into two sets: low and high eccentricity, however circularity, can be divided, into three sets: low, medium and high circularity.



(a) (b)

Fig. 5. Representation in the Plan of (a) Circularity Values and (b) Eccentricity Values of Reference Data Set of Shapes

In conclusion, the features convexity, concavity and circularity are fuzzified into three fuzzy sets of three fuzzy values: low, medium and high. However, the feature eccentricity is fuzzified into two fuzzy sets of two fuzzy values: low and high. Each fuzzy set is represented by a fuzzy membership function.
We choose trapezoidal function to represent fuzzy sets. Figure 6 [12] below represents the fuzzy sets "low, medium and high" for the feature convexity.

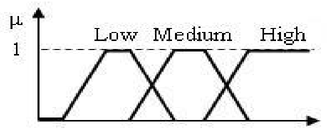


Fig. 6. Convexity Membership Functions

The crisp values of the features of shapes are fuzzified by computing the membership degrees to correspondent functions. After fuzzification, each shape is defined by fuzzy values as:

- two fuzzy values for eccentricity (i.e. membership degree to the fuzzy set "low" and membership degree to the fuzzy set "high")
- three fuzzy values for circularity
- three fuzzy vectors for concavity (i.e. vectors corresponding to membership degrees of ordinates of concavity points to fuzzy sets "low", "medium" and "high" respectively).
- three fuzzy vectors for convexity

C. Application of Fuzzy Similarity Measures to Classification of Shapes

Our objective is to compare two shapes having fuzzified attributes according to fuzzy sets of features. Consider shapes A and B described by M linguistic variables v_i (i.e. eccentricity, circularity, concavity and convexity) and for each linguistic variable v_i, k linguistic values L_k^i (i.e. low, medium, high) are defined. Each linguistic value L_k^i is represented by a fuzzy set with a membership function $\mu_{L_k^i}$. The similarity of two shapes A and B is computed on two steps:

- Compute similarity $S_{v_i}(A, B)$ according to each variable v_i
- Compute the similarity $S(A, B)$ according to all linguistic variables. This is done by computing the mean of all similarity degrees $S_{v_i}(A, B)$ using the following formula.

$$S(A, B) = \sum_{i=1}^{f} \frac{S_{v_i}(A, B)}{f} \tag{26}$$

Where f is the number of fuzzy linguistic variables.
We note that for the linguistic variables convexity and concavity $S_{v_i}(A, B)$ is computed by this formula:

$$S_{v_i}(A, B) = \frac{\sum_{k=1}^{n} S_{L_k^i}}{n}$$

with n is the number of fuzzy values of the i^{th} linguistic variable.

D. Experimentation and Results

In the classification process, test shapes are compared to reference shapes according to K-Nearest Neighbors algorithm. Table II presents the obtained results after applying similarity measures from literature. We note that 10-best represents the rate of classified shapes in 10-Nearest Neighbors.

TABLE II
RESULTS OF SHAPE CLASSIFICATION USING FUZZY SIMILARITY MEASURES

FSM	10-Best	Error rate (%)
S_3(10)	89.23	10.77
S_1(4)	88.95	11.05
D_E(20)	88.4	11.6
M(1)	88.12	11.88
S(15), S_5(6), S_4(5)	87.85	12.15
S_6(7), S_9(11), L(14)	87.57	12.43
S_L (24), d(22), d_2(25)	87.29	12.7
W(18), S_2(19)	87.29	12.7
S_{11}(13)	87.01	12.98
S_G (24)	86.19	13.81
S_{GO} (24)	85.91	14.09
T (2)	82.04	17.96
S_8 (9), S_{10} (12)	70.17	29.83
S_7 (8)	69.89	30.11
P (3)	68.78	31.22
S_{D_L} (23)	65.75	34.25

The results of classification in table II show that the first, the second and the third rank of similarity measures results are very closed and are obtained by S_3, S_1 and D_E respectively. The forth rank is obtained by the measure M followed by the results of the measures S, S_4 and S_5. In the sixth rank, the measures S_6, S_9, L produce very close results to their previous and their successors S_L, d, d_2, W, S_2 and S_{11}. The results decrease slightly with the measures S_G, S_{Go} and decrease more with the measure T. With the measures S_8, S_{10}, S_7, P and S_{D_L} the results decrease more and more and reach the last ranks.

VII. MOSAIC RECOGNITION

A. Data Set Images Presentation

We use images from the national library of Tunisia [1] which contain colored and gray scaled images with important historical value. We have only nineteen images. So, we experiment fuzzy similarity on two images of them.

B. Features Description

The features of images concern the shapes which compose them. So, the edges of some shapes in the image are described by the extended CSS descriptor defined in section VI by [29]. The features are fuzzified according triangular memberships ([11], [29]), the process of fuzzification is the same presented in figure 1. Despite that the features are the same used for shape classification presented in the precedent section, the application of fuzzy similarity measures to mosaic recognition is a different application and the experimentation results should give new ideas about FSMs.

[1]http://www.bibliotheque.nat.tn

C. Experimentation and Results

The images are compared according to their shapes. For each shape contained in the test images, a number of reference images are found, we present an example in figure 7 which contains some shapes. In this image, only the animal and the person are described with fuzzy extended CSS descriptors.

Fig. 7. An Image Containing some Shapes as a Tree, a Person and an Animal

The program matches images using the fuzzy features of animal edge. The expected images are all those containing animal shapes. The obtained images using the fuzzy similarity S (15) are presented in figure 8.

Fig. 8. Images Found while Comparing the Animal Shape in Figure 7.

Note that the resulting images shown in figure 8 contain animals.
In order to compare FSMs from literature, we apply them, to two images test: image1 and image2. We present the results in the tables IV, V and VI. Each table presents the results for one image compared according to one of its shapes. The comparison is done according to two shapes for image1 (i.e. shape1 and shape2) and to one shape for image2. The number of reference images containing shapes similar to shape1 in image1, is three and the number of shapes similar to shape1 in the three images is 10 (i.e., the first image contains 6 shapes, the second contains 3 and the third image contains 1 shape). So, the results of matching between shape 1 and reference images must be 10. The table below resumes for each test shape/image the number of reference images.

TABLE III
NUMBER OF SHAPES FOR IMAGES REFERENCE EXPECTED AS RESULTS TO TEST IMAGES

Image1	Shape1	Reference Images Number	3
		Shapes Number	10
		Expected Images Number	10
	Shape2	Reference Images Number	4
		Shapes Number	8
		Expected Images Number	8
Image2	Shape	Reference Images Number	3
		Shapes Number	4
		Expected Images Number	4

In the following, results are obtained according to KNN classifier. We choose k=10 for the shape1 of image1, k=8 for shape2 of image1 and k=4 for the shape in image2. The values of k correspond, respectively, to the number of reference images containing similar shapes of the test images.

TABLE IV
MOSAICS OBTAINED WITH FUZZY SIMILARITY MEASURES FOR IMAGE1, SHAPE1

FSMs	10-best (%)
W, S_4, S_8, d, S_{10}, D_E, d_2	90
S_1, S_{11}	80
S, S_2, S_7, S_9, S_{D_L}	70
S_6	60
S_L, S_3	50
M , L, S_5	40
S_G	10
T, P, S_{GO}	0

TABLE V
MOSAICS OBTAINED WITH FUZZY SIMILARITY MEASURES FOR IMAGE1, SHAPE2

FSMs	8-best (%)
W, S_4	50
S_1, S_2, S_8, S_{11}, D_E, d,d_2, S_3	37.5
M, S, L, S_{10}, S_5	25
S_{D_L}, S_L	12.5
T, P, S_{GO}, S_G, S_9, S_6, S_7	0

TABLE VI
MOSAIC OBTAINED WITH FUZZY SIMILARITY MEASURES FOR IMAGE2, SHAPE

FSMs	4-best (%)
S, S_{D_L}, S_1, S_9, S_2, S_8, P	50
S_{11}, D_E, d, d_2, S_L	50
S_3, W, S_4, M, S, L, S_{10}	50
T, S_{GO}, S_G	0

In table VI, there are two results. All measures produce 50% except measures T, S_{GO} and S_G which produce 0% as recognition rate in 4-best . The tables IV and V present different results for the comparisons between the mosaics based on two shapes in the test image. In the two tables, W and S_4 are in first rank and S_1 and S_{11} are in second rank. The measures S_8, d, D_E, d_2 are in first rank in the table IV and in second rank in the table V. However, the measures S_2, S_3 are in second rank in the the table V and in third and fifth rank, respectively, in the table IV. The measure S_{10} is in the the first rank in table IV and the third rank in table V. Thus, we can deduce that best results in the tables (IV, V) are given by the measures: S_8, d, D_E, d_2, W, S_4, d.

Results decrease with measures S, S_{D_L} which are in third rank in table IV and in forth and fifth rank, respectively in table V. S_6 is in forth rank in table IV with modest results and in last rank in table V with 0% recogntion rate. S_9 and S_7 are in third rank in table IV with 70% recogntion rate and in fifth rank in table V with 0% recogntion rate. Modest results are given by S_L which is in fifth rank in table IV and in forth rank in table V followed by the results of M, L, S_5 which are in third rank in table V and in sixth rank in table IV. The worst results are obtained with the measures T, S_{GO} and P, S_G with 0% recognition rate in the three Image/shape comparisons, except the results of S_G for image1, shape1 with a recognition rate of 10% in table IV. Also S_9, S_6, S_7 produce 0% of recognition rate for image1, shape2 in table V.

In this application there is a few number of test images. So, the decision for better results is not clear, but indicates that the measures d_2, W, S_4, S_1, S_{11}, d, D_E, S_8 produce best results. The results of measures S_1, D_E, S_4, d, d_2, W, S_{11} confirm those found in shape classification in section VI-D. The results produced with P, T, S_G, S_{GO}, are in last ranks in shapes classification but better than those obtained for mosaic recognition.

VIII. Arabic Sentences Recognition

A. Data Set Sentences Presentation

We use 6537 images of 824 handwritten Tunisian town/village names extracted from the IFN/ENIT data set [30]. The images of data set are written by different writers and undergo some preprocessing like noise reduction but are not normalized. In this data set, every image is described by some information like the number of characters and the number of connected compounds. We divide data set into two data sets: the first serves as reference data set constituted of 4357 word images and the second serves as test data set constituted of 2180 word images. The recognition process is done using fuzzy sets on two steps:

- the description of images by fuzzy features.
- the comparison between test data set and reference data set using a fuzzy similarity measure

In the following subsections, we describe these processes with more details.

B. Features Extraction

We extract features based on some information existing in the IFN/ENIT data set [30] and without normalization of images. The data set is described with these information:

- the number of sentence characters
- the number of sentence connected compounds
- the coordinates of the top line and those of the base line of the word
- the description of each sentence connected compound edge with Freeman chain code [31]

We are interested in extracting features from word connected compounds, so we delete diacritic signs from images because they can be misplaced and cause noise for recognition. The extracted features from the word connected compounds are:

- sum of black pixel distances from the base or the top lines
- higher black pixel coordinates
- code frequencies of freeman chain code

a) Sum of Black Pixel Distances From the Base or the Top Lines: We consider the top line and the base line of an image as lines in a plan having the equation: $y = ax + b$. Each pixel in the image is considered as a point in the plane. So, the distance of a pixel to the baseline of the image or to the top line of the image is computed as mathematical distance of a point to a line. For each black pixel, in the edge of connected compounds, three mathematical distances are computed as:

- d_1: the sum of distances of each pixel from the top line if the pixel is upper this last (figure 9. A)
- d_2: the sum of distances of each pixel from the baseline if the pixel is under top line and upper the baseline (figure 9. B)
- d_3: the sum of distances of each pixel from the baseline if the pixel is under this last (figure 9. C).

Each distance is divided by the sum of all black pixels of connected compound edge denoted s. We obtained: $d_1 = \frac{d_1}{s}$, $d_2 = \frac{d_2}{s}$ and $d_3 = \frac{d_3}{s}$.

Figure 9 shows two words of a sentence, the first word is constituted with two connected compounds and the second is composed by four connected compounds.

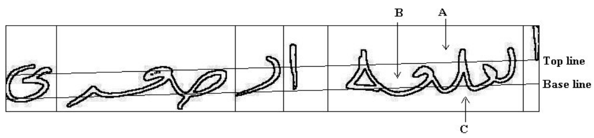


Fig. 9. Connected Compounds of Two Words of a Sentence

b) Higher Black Pixel Coordinates: Finding the coordinates of the higher black pixel in a connected compound can indicate the presence of a stroke. This information can be determinative to differentiate between connected compounds and for consequence, it can differentiate between words. The searched pixel can be in the beginning, in the middle or in the end of the connected compound.

c) Direction Frequency of Freeman Chain Code: Each connected compound is described by a chain code of freeman which is constituted by a succession of numbers from zero to seven. The numbers indicate directions of a pixel constructing the connected compound edge. This directions can be vertical, horizontal or representing a curvature. We compute the frequency of each number in the freeman chain code as N/T where N is the number of the code in freeman chain code and T is the total number of codes. So, we obtain eight informations. It is obvious that all obtained informations are in the interval $[0, 1]$ (i.e. N is less than T).

C. Features Fuzzification

The crisp features presented in the precedent section are fuzzified in order to be represented with fuzzy sets. The process of fuzzification is presented in figure 1.

- The sum of pixel distances presented in the section VIII-B0a are fuzzified by computing the membership degree of each of them to trapezoidal functions.
- The ordinates of the higher black pixel found in the section VIII-B0b are fuzzified by computing their

membership degrees according to triangular fuzzy sets. For each connected compound, we retain three degrees of belongingness. The first, represents the membership degrees of a pixel in the beginning of connected compound. The second, represents the membership degree of a pixel in the middle of connected compound and the last represents the membership degree of a pixel at the end of connected compound.

- We are interested to the pixels indicating the curvature direction because of their importance. In Freeman chain code, the codes of pixel directions are from 0 to 7. The codes 1, 3, 5 and 7 indicates the pixel belongingness to a curvature. Thus, we consider that the frequencies of these codes (see Sect. VIII-B0c) represent their membership degrees to indicate curvature. The codes 0, 2, 4 and 6 indicates the pixel belongingness to a vertical or horizontal direction. Therefore their frequencies (see Sect. VIII-B0c) represent their membership degrees to indicate horizontal or vertical positions.

D. Experimentation and Results

Table VII. shows the results of Arabic sentences recognition with applied fuzzy similarity and distance measures. We recall that 10-best is the rate of found sentences in the first 10 positions to be similar to the test sentence. The results of

TABLE VII
ARABIC SENTENCE RECOGNITION RESULTS OBTAINED WITH FUZZY SIMILARITY MEASURES

FSMs	10-best (%)	Error rate (%)
S_1 (4)	91.10	8.89
S_3(10)	90.92	9.08
S_{D_L}(23)	90.6	9.4
d_2 (25)	90.28	9.72
L (14), S_L (24)	90.23	9.77
d (22)	89.82	10.18
S_4 (5)	89.68	10.32
S_9 (11)	89.59	10.41
S_6 (7)	89.59	10.41
W (18), S_2 (19)	89.59	10.41
S_7 (8)	89.31	10.69
D_E(20)	87.47	12.52
S_{11}(13)	85.55	14.45
S_{10}(12)	84.68	15.32
S_8(9)	84.59	15.41
S(15)	84.27	15.73
S_5(6)	75.6	24.4
M (1)	62.93	37.06
P (3)	49.95	50.05
S_G (24)	2.06	97.93
T (2)	1.83	98.17
S_{GO} (24)	0	100

the measures S_3, S_1, S_{D_L}, d_2 are respectively in first, second, third and fourth rank with very close results and a difference varying from 0.18 to 0.32. The results of the measures L, S_L are in fifth rank and are very closed to those of d_2 with a difference in results of 0.05. The results of the measure d are in sixth rank having a difference of 0.41 from previous results, followed by the results of S_4 with a difference 0.09 in results, then, S_9, S_6, W and S_2 with a difference in results of 0.23. The results of the measure S_7 are in eighth rank with a difference of 0.27 from the previous results. The results decrease using the measure D_E with a difference of 1.84 and using the measure S_4 with a difference of 1.74, from their previous. Results continue decreasing with the measures S_{11}, S_{10}, S_8, S with bit differences. The decrease in results becomes more important with the measure S_5 and M which present differences of 8.67 and 21.34, respectively, from previous results. The measure P produces low results followed by the measures S_G, T and S_{GO} which produce the worst results.

IX. Comments on obtained results with application of the similarity measures on the three data sets

In this section, we develop the relationship between the obtained results and the similarity and distance properties' validation. Our intention is to examine the influence of the latter on the results obtained in three applications with three data sets. The measures producing low results for each application are in the following:

- classification of shapes : S_8, S_{10}, S_7, P and S_{D_L}.
- Recognition of mosaics: P, S_G and S_{GO} for all tests images, T for two tests images and S_{D_L}, S_L, S_9, S_6, S_7 for one test image among three.
- Recognition of handwritten Arabic sentences: S_5, M, T, P, S_G and S_{GO}.

In conclusion, for the three applications the results of the measures S_5, M, P, S_G, T, S_{GO}, S_{D_L}, S_9, S_6, S_{10} and S_8 did not produce good results. Let us examine their properties:

- T, S_7, S_8, S_{GO}, S_{10} do not verify the properties 3 and 4
- P, S_{D_L} do not satisfy the property 4
- S_G, S_L, M, S_6, S_5 validates all properties

These results show that:

- The properties 3 and 4 must be satisfied by the measures, otherwise the results of classification can be not good (i.e., S_{11} produced acceptable results in the three applications).
- The validation of the property 4 by measures is without effect in results in general. Most measures which does not validate this property produce good results such as S_3.

We can conclude that the satisfaction of the third property by similarity and distance measures is essential to obtain acceptable results for classification. In addition, the satisfaction of all properties by the similarity measures does not imply that they produce good results (i.e, low results in all applications are obtained with S_G which satisfies all properties).

X. Conclusion

We presented and discussed properties of similarity and distance measures. Thus, we provided common existing properties for distance and similarity measures and we validated measures from literature according to these common properties. We presented three data sets described with fuzzy features and we applied FSMs for classification and recognition of test data sets. These applications permit us to conclude that good results can be produced using measures S_3, S_1, d_2 and low results can be obtained using measures T, S_7, S_8, S_{10}, S_{GO}, P and S_G despite that some of them can produce good results in some cases.

This study, shows the importance of validation of FSMs to the modified properties (i.e. third property) and classifies measures according to their results in three applications. Therefore, the obtained results can simplify the choice of similarity measures, from numerous existing measures in literature, for any research topic.

REFERENCES

[1] S. M. Chen, M. S. Yeh, and P. Y. Hsiao, "A comparison of similarity measures of fuzzy values," *Fuzzy sets and systems*, vol. 72, pp. 79–89, 1995.

[2] J. Fan and W. Xie, "Some notes on similarity measures and proximity measures," *Fuzzy sets and systems*, vol. 101, pp. 403–412, 1999.

[3] Y. S. Hyung, L. K. Song and K. M. Lee, "Similarity measures between fuzzy sets and between elements," *Fuzzy sets and system*, vol. 62, pp. 291–293, 1994.

[4] C. P. Pappis and I. Karacapilidis, "A comparative assessment of measures of similarity of fuzzy values," *Fuzzy sets and systems*, vol. 56, pp. 171–174, 1993.

[5] X. Z. Wang, B. D. Baets, and E. Kerre, "A comparative study of similarity measures," *Fuzzy sets and systems*, vol. 73, pp. 259–268, 1995.

[6] I. Bloch, "On fuzzy distances and their use in image processing under imprecision," *Pattern Recognition*, vol. 32, pp. 1873–1895, 1999.

[7] D. V. d. Weken, M. Nachtegael, W. V. De, S. Schulte, and E. E. Kerre, "A survey on the use and the construction of fuzzy similirity measures in image processing," in *CIMSA 2005-IEEE International Conference on Computational Intelligence for Measurments Systems and Applications*, 2005.

[8] D. V. d. Weken, M. Nachtegael, and E. E. Kerre, "Using similarity measures and homogeneity for the comparison of images," *Image vision computing*, vol. 22, pp. 695–702, 2004.

[9] D.-G. Wang, Y.-P. Meng, and H.-X. Li, "A fuzzy similarity inference method for fuzzy reasoning," *Computers and mathematics with applications*, vol. 56, pp. 2445–2454, 2008.

[10] T. Gadi, R. Benslimane, M. Daoudi, and S. Matusiak, "Fuzzy similarity measure for shape retrieval," in *Vision interface 99, Trois-rivière, Canada*, 1999.

[11] L. Baccour, W. Maghrebi, S. Kanoun, and A. M. Alimi, "A comparative study of a fuzzy similarity measure aggregated with fuzzy implications applied to shape recognition," in *proceedings of the 6th Mexican International Conference on Artificial Intlligence, Proc. of MICAI 2007, Aguascalientes, Mexico*, 2007.

[12] L. Baccour and A. M. Alimi, "Applications and comparisons of fuzzy similarity measures," in *IEEE International Conference on Fuzzy Systems, FUZZ-IEEE 2010, Barcelone, July 18-23*, 2010.

[13] L. Baccour, A. M. Alimi, and R. I. John, "Similarity measures for intuitionistic fuzzy sets : State of the art," *Journal of Intelligent & Fuzzy Systems*, vol. 24 (1), pp. 37–49, 2013.

[14] R. Buse, Z. Liu, and J. Bezdek, "Word recognition using fuzzy logic," *Ieee transactions on fuzzy systems*, vol. 10 (1), pp. 65–76, February 2002.

[15] M. Ismail and S. Gad, "Off-line arabic signature recognition and verification," *Pattern Recognition*, vol. 33, pp. 1727–1740, 2000.

[16] G. Lu and R. Liu, "Discrimination for fuzzy sets related to ntv metric," *IAENG International Journal of Applied Mathematics*, vol. 43 (4), pp. 204–208, 2013.

[17] A. K. Abd-Elaal, H. A. Hefny, and A. H. Abd-Elwahab, "Forecasting of egypt wheat imports using multivariate fuzzy time series model based on fuzzy clustering," *IAENG International Journal of Computer Science*, vol. 40 (4), pp. 230–237, 2013.

[18] H. Hedayati and A. Ebrahimian, "Abnormal s-fuzzy subpolygroups," *IAENG International Journal of Applied Mathematics*, vol. 42 (1), pp. 34–39, 2013.

[19] R. Pèrez-Aguila, "Enhancing brain tissue segmentation and image classification via 1d kohonen networks and discrete compactness: An experimental study," *Engineering Letters*, vol. 21 (4), pp. 171–180, 2013.

[20] L. Zadeh, "Fuzzy sets," *Inform. Control*, vol. 8, pp. 338–356, 1965.

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

[22] W.-J. Wang, "New similarity measures on fuzzy sets and on elements," *Fuzzy sets and systems*, vol. 85, pp. 305–309, 1997.

[23] S. Raha, A. Hossain, and G. S., "Similarity based approximate reasoning: fuzzy control," *Journal of applied logigic*, vol. 6, pp. 47–71, 2008.

[24] R. Zwick, E. Carlstein, and V. Budescu, D., "Measures of similarity among fuzzy concepts: a comparative analysis," *Approximate Reasoning*, vol. 1, pp. 221–242, 1987.

[25] J. Fan and W. Xie, "Distance measure and induced fuzzy entropy," *Fuzzy Sets and Systems*, vol. 104, pp. 305–314, 1999.

[26] D. Bhandari and N. R. Pal, "Some new information measure of fuzzy sets," *Inform. science*, vol. 67, pp. 209–228, 1993.

[27] F. Mokhtarian, S. Abbasi, and J. Kittler, "Efficient and robust retrieval by shape through curvature scale space," in *Proceedings of the first International Workshop on Image Databases and Multimedia Search, August*, 1996, pp. 35–42.

[28] S. Kopf, T. Haenselmann, and W. Effelsberg, "Shape-based posture recognition in videos," *Proceedings of Electronic Image, January*, vol. 5682, pp. 114–124, 2005.

[29] W. Maghrebi, L. Baccour, A. M. Khabou, and A. M. Alimi, "An indexing and retrieval system of historic art images based on fuzzy shape similarity," in *proceedings of the 6th Mexican International Conference on Artificial Intelligence, MICAI '2007*, 2007, pp. 623–633.

[30] M. Pechwitz and al., "Ifn/enit – data set of handwritten arabic words," in *proceedings of CIFED 2002, Hammamat, Tunisia*, 2002, pp. 129–136.

[31] H. Freeman, "Computer processing of line-drawing images," *ACM Computing Surveys*, vol. 6 (1), pp. 57–97, 1974.