A Fuzzy Approach to Performance Evaluation of Edge Detectors

A. S. OSIPOV
Scientific Research Institute for System Studies, Russian Academy of Sciences
(NIISI RAN)
Nakhimovskii pr. 36-1, Moscow, 117218
RUSSIA

Abstract: - The system PICASSO 2 represents the latest version of software package, designed for a comparative evaluation of image processing algorithms. In this paper we discuss the part of the system which evaluates edge detectors and consider its fuzzy extension. Namely, we introduce the ground truth edge maps defined in the fuzzy way and complete the system with several fuzzy similarity measures. The proposed approach allows one to evaluate both fuzzy and non-fuzzy edge detection algorithms.

Key-Words: - Image processing, edge detection, performance metric, fuzzy ground truth, fuzzy similarity measure, membership function, image feature points.

1 Introduction

During four decades of computer vision research a lot of different methods were proposed. The wide selection of methods currently available provokes much interest in the development of techniques for choosing the most appropriate one for individual cases. The development of the methods which provide a comparative assessment of image processing algorithms is therefore becoming very topical [1-2]. At the same time, a growing amount of algorithms handling such tasks as edge detection, image restoration, boundary improvement and texture analysis relies on fuzzy set theory (see e. g. [3-5] and references thereafter). In many situations the application of these algorithms is justified. For instance, as pointed out in [6] “gray level images are inherently fuzzy in nature due to the uncertainty that exists in locating the exact position of the boundary which separates the object from background”. Obviously, the comparative evaluation technique must be able to handle these fuzzy algorithms. Also, the technique itself must include fuzzy elements.

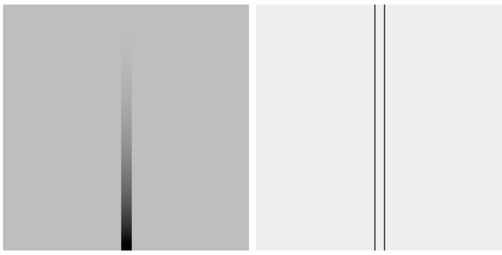
In order to obtain a tool for testing and evaluation of the image processing methods, we are developing the software system named PICASSO (PICture Algorithms Study SOftware). Originally it was designed to compare various edge detection algorithms on a set of artificial 2D images [7]. It exploits the so-called empirical discrepancy evaluation methods which use a ground truth (or reference image) – an ideal edge map for a given test image. The new version of the system named PICASSO 2 [8] evaluates a wider range of image processing methods. Also the testing technique has been improved. For example, nowadays some well known edge detectors (such that Sobel, Canny, etc.) have many software realizations. The testing approach implemented in PICASSO 2 can help the practical user to find out which realization is the most suitable for his practical needs (or at least to sort out the erroneous realizations).The main goal of our further research is to create an adaptive system for real image segmentation on the basis of PICASSO.

In the present paper, we analyze the part of our system which performs the evaluation of edge detectors, and consider its fuzzy extension. Namely, in addition to the previously used ground truth edge maps, we introduce a set of reference images, defined in a fuzzy way. For such images the grade of membership of each reference pixel to the edge class is given. To compare the outputs of different edge detectors (edge maps) with these fuzzy ground truths, we use the fuzzy similarity measures introduced in [9]. The advantage of these measures is that they allow one to evaluate the outputs of non-fuzzy edge detectors (like e. g. that of Canny) and to compare them with the outputs of their fuzzy counterparts.

The paper is organized as follows. In Section 2 we briefly describe the main features of PICASSO 2. We also discuss the advantages and disadvantages of the performance metrics implemented into the system. Section 3 contains a thorough description of the proposed approach with some examples of its practical application. Finally, Section 4 concludes the discussion.

2 PICASSO 2 – Basic Features and Performance Metrics Used

The original version of PICASSO system has been described in [7] and as mentioned above, PICASSO 2 represents its further extension. Note that the core feature of both PICASSO and PICASSO 2 is modeling of typical situations in image processing. We have worked out a set of synthetic grayscale images, as well as a set of corresponding ground truths, forming the image database of PICASSO 2. These synthetic images simulate a collection of situations, which are difficult in some sense for the image processing methods (see Fig. 1 as an example). Also the system includes the special image editor, software implementation for the methods tested, noise generators, filling templates for background and objects.



a) b)

Fig.1. a) Degenerating Ridge b) Its ground truth

As proposed by Canny [10], an edge detector should be considered 'good' if it exhibits good detection (low probability of failing to detect an edge and low probability of incorrectly labeling a background pixel as an edge) and good localization (points identified as edge pixels should be as close as possible to the centre of the true image). So nowadays various performance metrics (discrepancy measures) applied to the evaluation of edge detectors are often divided into two classes: detection performance ('statistical') measures and localization performance ('distance') measures (see e. g. [11]). In PICASSO 2 several measures from the both classes are implemented. Namely, let X denote the pixel raster, assumed to be a finite set. Let A be the ground truth image (edge) and a B the putative or "estimated" image. Then define the type I error rate [11] by

$$\alpha(A,B) = \frac{n(B \setminus A)}{n(X \setminus A)},$$

the type II error rate

$$\beta(A,B) = \frac{n(A \setminus B)}{n(A)},$$

and the overall misclassification rate

$$\varepsilon(A,B) = \frac{n(A \Delta B)}{n(X)},$$

where $n(S)$ = number of pixels in S, Δ denotes set symmetric difference. These are typical examples of statistical measures. Among the distance measures we consider, the Pratt's figure of merit FOM (where a scailing constant is usually set to 1/9) and the Hausdorff distance.

We also used in [7] the following couple of statistical measures: Sensitivity and Specificity (denoted by *Se* and *Sp* respectively)

$$Se = \frac{n(B \cap A)}{n(A)}; \qquad Sp = \frac{n(B \cap A)}{n(B)}.$$

These two measures represent the simplest versions of Producer's and, respectively, User's accuracy measures (see e. g. [13]). Sensitivity and Specificity are indeed quite similar to the above statistical measures, e. g. $Se=1-\beta$. At the same time, in some complicated cases (including a well known Peli-Malah example [11]) they give more realistic results than $\alpha, \beta, \varepsilon$ and FOM.

As mentioned in the introduction, there exists uncertainty in locating the border of a grayscale image (as e. g. on Fig.1 a)). For our evaluation method it means that the corresponding ground truth (like e. g. on Fig.1 b) which has one pixel width) represents only one (out of a several) possible versions of the "true edge map", and the other versions should be taken into consideration somehow. In our system this problem is partly solved by adding the Distance Threshold parameter to the statistical measures we use. This parameter specifies the minimal allowed distance between a detected edge pixel and a corresponding one on the ground truth. For example, if the Distance Threshold parameter is set to one, and we compare any other version of the Degenerating Ridge's ground truth with the image shown on Fig.1 b), the results will be the same as for the Fig. 1 b) compared with itself. Also, changing the threshold value allows the user to study the localization performance as well.

We cannot, however, in all cases rely on the results produced by our metrics. As an example, consider the reference image – one pixel wide step (Fig. 2 a)) and assume that we have the outputs of three different edge detectors given on Figs. 2) b) – d). Fig. 2 b) represents a similar step in the same location which is two pixel wide. The step on Fig. 2 c) is shifted to the right on one pixel, and Fig. 2) d) represents two parallel steps located between the original step. Comparing these three images with Fig. 2 a) and setting the Distance Threshold to one we obtain the same values of our statistical measures in all three cases. Only the values of FOM are different: 0.95 for Fig. 2 b) and 0.9 for the other two cases, thus giving the advantage to the thick step. But we are still unable to make a preference between the Figs. 2 c) and d) without the use of human eyes!

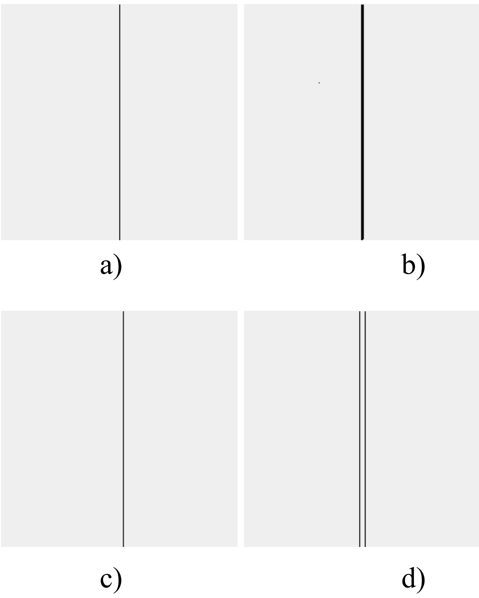


Fig.2. a) Ground truth - step b) Thick step c) Shifted step d) Two steps

In view of the above, we can make the following conclusions. First, because of uncertainty in locating the exact position of a reference image and related problem with the use of such images for empirical evaluation, an extension of a “ground truth” concept seems desirable. Second, it is widely agreed that in a considerable amount of cases the metrics used for evaluation of edge images give inadequate results (see e. g. [12]). Our practical results also show that the design of performance metrics still remains an important task for comparative study of edge detectors.

3 A Fuzzy Extension

As we see, the method of empirical discrepancy evaluation is based on two concepts: reference (ground truth) image and discrepancy measure (performance metric). To cope with the above problems of edge detection performance evaluation, we completed the collection of PICASSO’s reference images with a number of ground truths defined in a fuzzy way. To evaluate the fuzzy edge images, we also embedded into our system a few performance metrics based upon the concept of fuzzy similarity [9].

First of all, we recall some basic notions of the fuzzy sets theory. Let X be a non-empty set (e. g. the pixel raster). A fuzzy set C in X (called the support set) is a pair $<X, f_C>$, where f_C is a mapping from X into [0, 1]. The value $f_C(x)$ at a point $x \in X$ is called the grade of membership of x to C, and the function f_C is called the fuzzy set membership function (FSMF). We denote the set of all fuzzy sets in X by $[0, 1]^X$. For A and B from $[0, 1]^X$ the fuzzy set inclusion

$A \subset B$ means that $f_A(x) \le f_B(x)$ for all $x \in X$.

A fuzzy similarity measure is a mapping

$s : [0, 1]^X \times [0, 1]^X \to [0,1]$ assigning two fuzzy sets $A, B \in [0, 1]^X$ a degree of similarity $s(A,B) \in [0, 1]$ which is subject to the conditions

- $s(A,A) = 1$ for every fuzzy set A
- $s(A,B) = s(B,A)$ for all fuzzy sets A and B
- $s(A,C) \le s(A,B) \wedge s(B,C)$ whenever $A \subset B \subset C$,

where $p \wedge q$ denotes the minimum of p and q; the maximum of p and q is denoted by $p \vee q$. Important in the sequel will be the following examples (implemented in the current version of PICASSO):

$$s_1(A,B) = \frac{\sum_x (f_A(x) \wedge f_B(x))}{\sum_x (f_A(x) \vee f_B(x))}, \qquad (1)$$

$$s_2(A,B) = \frac{2\sum_x (f_A(x) \wedge f_B(x))}{\sum_x (f_A(x) + f_B(x))} \qquad (2)$$

From the equalities $(p \vee q) - (p \wedge q) = | p-q |$ and $p + q = 2(p \wedge q) + | p-q |$ we obtain

$$s_1(A,B) = 1 - \frac{\sum_x | f_A(x) - f_B(x) |}{\sum_x (f_A(x) \vee f_B(x))},$$

and

$$s_2(A,B) = 1 - \frac{\sum_x | f_A(x) - f_B(x) |}{\sum_x (f_A(x) + f_B(x))}.$$

Note that ordinary (crisp) subsets M of X are covered by the fuzzy set approach if we view them as standard characteristic functions $1_M : X \to [0,1]$. Thus the measures s_1 and s_2 can also be applied when one or both of A and B are crisp sets. As shown in [9], for the crisp A and B viewed as characteristic functions, s_1 and s_2 coincide with Short’s and, respectively, Hellden’s accuracy measures originally defined only for crisp sets, so one may consider s_1 and s_2 as a generalization of these accuracy measures.

As to the fuzzy ground truths, there are no general rules for their construction. For example, as noted in [9], for remote sensing applications “it is a big and possibly up to now unsolved problem to obtain reliable ground truth". We included the fuzzy ground truths in our system, aiming not only to have a tool for testing the fuzzy edge detectors, but also to make the existing evaluation procedure more profound. We believe that depending on the specific feature being tested, different fuzzy ground truths corresponding to the same test image can be used. A fuzzy edge detector (see e. g. [4-6]) typically differentiates pixels into the following six classes: four edge classes, a background class and a speckle edge class (a speckle is a noisy pixel). Each class has its own membership function. The edge classes correspond to the four directions in which edges may appear. For our tests, we selected the images from our database where the edges of only one or two classes are present with no noise added. Thus the corresponding fuzzy ground truths must have two or three reference classes (for edges and background). To compare sets which contain several fuzzy classes a couple of overall accuracy measures based upon the measures defined by (1)-(2) has been offered in [9]. Namely, let $B = \{B_1, B_2,...,B_N\}$ be a fuzzy classification and $A = \{A_1, A_2,...,A_N\}$ be a system of fuzzy reference classes . Then the overall accuracy measures are defined by the formula

$$OA_i(A,B) = s_i([A_1,\cdots,A_N],[B_1,\cdots B_N]), \quad i=1,2.$$

For example, for $i=1$ we have

$$OA_1(A,B) = \frac{\sum_{k=1}^{N}\sum_{x}(f_{Ak}(x) \wedge f_{Bk}(x))}{\sum_{k=1}^{N}\sum_{x}(f_{Ak}(x) \vee f_{Bk}(x))}.$$

It was shown in [9] that OA_2 generalize the standard overall accuracy measure OA (see e. g. [13]) in the sense that for a crisp classification and a system of crisp reference areas viewed as characteristic functions we get $OA_2=OA$.

In order to check how these fuzzy elements affect the evaluation results, we apply them to study the non-fuzzy edge detectors previously considered in [7-8], where the performance metrics from the previous section were applied. Due to the limited paper length, only some important results will be given here. For example, take the 256 x 256 pixel test image – Degenerating Step and apply to it the detector of Heitger and that of Smith (also known as SUSAN) (see [8] and references thereafter). The corresponding results are shown on Fig. 3. As illustrated in the Fig. 3 a), the left side of Degenerating Step contains the background pixels (with gray-scale value equal to 190) and its right side contains the pixels of varying grey levels (from 190 in the upper part of the picture to 0 in its lower part). It seems natural to assume that the edges of only one class (vertical) are presented on this picture (see Fig. 3 b).

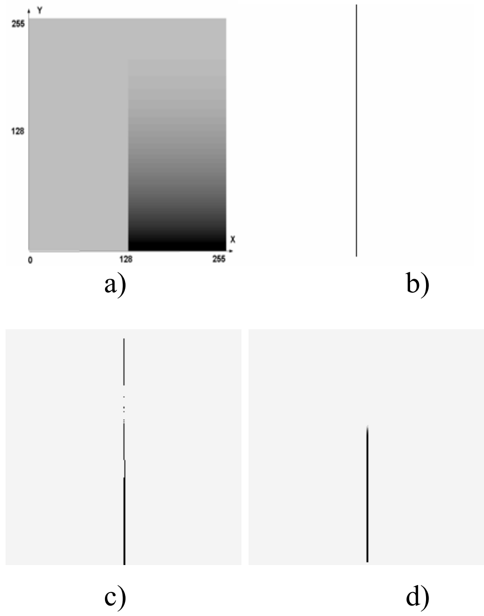


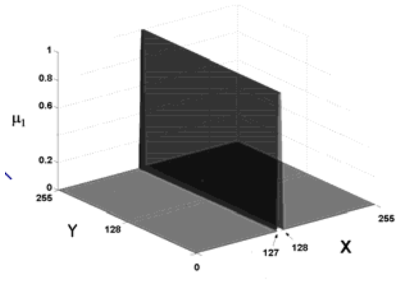
Fig.3. Results of processing Degenerating Step by Heitger and Smith algorithms: a) original picture b) ground truth c) Heitger edge map d) Smith edge map.

Comparing the Figs 3 c) and d) with the ground truth by using the non-fuzzy metrics from our measurement toolbox, we obtain the following results (with the Distance Threshold set to one).

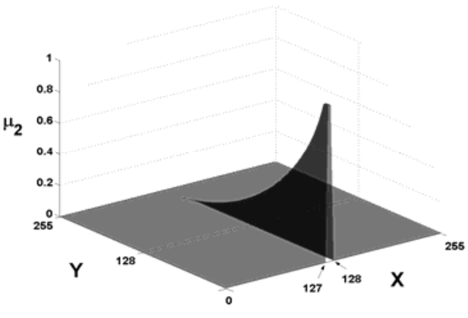
Pictures	*Se*	*Sp*	FOM	*H*
Heitger edge map (Fig. 3c))	0.867	1.00	0.962	10
Smith edge map (Fig 3 d))	0.589	1.00	0.949	105

Table 1: Performance evaluation of Heitger and Smith detectors applied to Degenerated Step by means of PICASSO 2 system (here *H* stands for Hausdorff distance).

As we see from the Table 1, none of these metrics produced better result for the Smith detector. Now, instead of the reference image shown on Fig. 3 b), we introduce two fuzzy ground truths G_1 and G_2, generated by functions μ_1 and μ_2 which determine the degree of membership for the (vertical) edge class (Fig. 4). The function μ_1 is a rectangular membership function giving the value of one to the pixels with x- coordinates 127 and 128. In other words, μ_1 is the characteristic function of the thick step shown on Fig. 2 b). The second function, μ_2, takes into account the relative edge strength of the pixels in Fig. 3 a) assigning the value of one to the pixels in the lower middle part of this picture.



a)



b)

Fig.4. Fuzzy ground truths for Degenerated Step : a) G_1 b) G_2

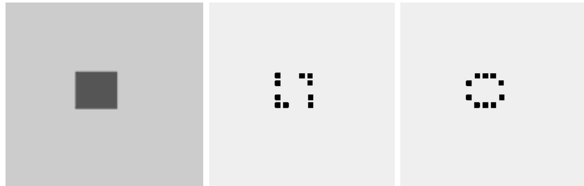
Considering the edge maps shown on Figs 3 c) and d) in terms of their characteristic functions, one may compare them with G_1 and G_2 using the fuzzy measures s_1 and s_2 defined by (1)-(2). Table 2 presents the results.

Edge detector	Ground truth G_1		Ground truth G_2	
	s_1	s_2	s_1	s_2
Heitger	0.602	0.7512	0.3547	0.5236
Smith	0.581	0.7326	0.5243	0.6853

Table 2: Performance evaluation of Heitger and Smith detectors applied to Degenerated Step. Fuzzy ground truths and fuzzy measures.

We see that the comparison with G_1 indicates the advantage of the Heitger edge detector (which confirms our previous results), whereas the comparison with G_2 shows the advantage of the Smith detector. One can explain this by the fact that the function μ_1 assigns the same membership values to the weak edge pixels as to the strong ones (and therefore G_1 can be used to test the detection of weak edges). At the same time, the values of μ_1 for edge pixels depend on their location (and strength), thus making it sensitive to the gaps in the edge map (as on Fig 3 c)).

In developing performance criteria for an edge detector, it is important to evaluate its ability to detect image feature points. For instance, the knowledge of these points is important for the edge linking procedure. Our fuzzy approach can be useful for such evaluation. To illustrate this, consider the test image – 5 x 5 pixel dark square and apply to the Canny edge detector and the detector of Rothwell (Fig. 4). As we see, all four corner points are detected by the first method and skipped by the second one.



a) b) c)

Fig.5. Results of processing 5 x 5 pixel square by Canny and Rothwell algorithms (enlarged images): a) original picture b) Canny edge map d) Rothwell edge map.

It is not hard to see, that for the corresponding non-fuzzy ground truth the results of its comparison with the Figs 5 b) and c) will be the same in terms of the metrics mentioned in the Table 1. On the Fig. 5 a) the edges of two classes (horizontal and vertical) are present. If we define two membership functions assigning the value of 0.8 to all edge pixels except for the corner ones (which belong to both classes), and for the corner pixels we set the value of 1, we obtain the fuzzy ground truth shown on Fig. 6.

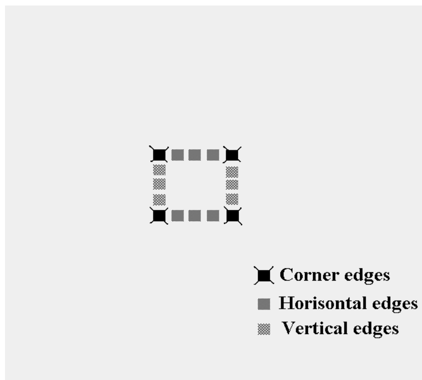


Fig.6. Fuzzy ground truth for 5 x 5 pixel square.

Comparing the Figs 5 b) and c) with the Fig. 6 by using the overall (with respect to the both edge classes) accuracy measures OA_1 and OA_2, we get the results presented in Table 3.

Edge detector	OA_1	OA_2
Canny	0.61	0.75
Rothwell	0.41	0.58

Table 3: Performance evaluation of Canny and Rothwell detectors applied to 5 x 5 pixel square. Fuzzy approach.

Both measures in this example indicated a clear preference for the Canny detector.

4 Conclusion

As mentioned above, one of the advantages of the proposed approach is its ability to evaluate both fuzzy and non-fuzzy algorithms (as a matter of fact, only some results for non-fuzzy edge detectors are presented in this paper). Also the idea to use several fuzzy ground truths corresponding to the same test image to study different features of the tested method deserves further consideration. Another important feature of our method is that in some cases it allows one to check the ability of an edge detector to find image feature points.

We also mentioned that the fuzzy similarity measures considered in the previous section represent a generalization of some non-fuzzy detection performance ('statistical') measures. These similarity measures, however, are not able to provide a proper evaluation of the localization performance. For example, if we use them to compare the ground truth shown on Fig. 3 b) with the vertical step having x – coordinate 129, we get zero in all cases. It contradicts to the common sense assumption that if we compare similar images slightly displaced with respect to each other, the error must be small. In recent years several fuzzy localization performance measures have been offered. In particular, they include some fuzzy versions of the Hausdorff metric [14]. Our next goals are to implement these measures in PICASSO and to complete the image database of our system with a collection of fuzzy ground truths, intended for testing various features of edge detection algorithms.

References:

[1] T. Pavlidis, Why Progress in Machine Vision is So Slow, *Pattern Recognition Letters*, Vol.13, No.4, 1992, pp. 221-225.

[2] R.M. Haralick, Performance Characterization in Computer Vision, *CVGIP: Image Understanding*, Vol. 60, No 2, 1994, pp. 245-249.

[3] S.K. Pal, D.D. Madjumadar, Fuzzy Mathematical Approach to Pattern Recognition, New York: Wiley, 1986.

[4] H. R. Tizhoosh, Fuzzy Image Processing : Introduction in Theory and Applications, Springer-Verlag, 1997 (in German).

[5] L. R. Liang, C.G. Looney, Cometitive fuzzy edge detection, *Applied Soft Computing,* Vol. 3, 2003, pp. 123-137.

[6] M. Banerjee, M.K. Kundu, Content Based Image Retrieval with Fuzzy Geometrical Features, *Proceedings of the IEEE International Conference on Fuzzy Systems,* 2003, pp. 932-937.

[7] I.V. Gribkov, P.P. Koltsov, N.V. Kotovich, A.A. Kravchenko, A.S. Koutsaev, V.K. Nikolaev, A.V. Zakharov, PICASSO – Edge Detectors Evaluating System Based on the Comprehensive Set of Artificial Images, *Proceedings of the 6th World Multiconference on Systemics, Cybernetics and Informatics*, Vol. 9, 2002, pp 88-93.

[8] I.V. Gribkov, P.P. Koltsov, N.V. Kotovich, A.A. Kravchenko, A.S. Koutsaev, A.S. Osipov, A.V. Zakharov, Empirical evaluation of Image Processing Methods Using PICASSO 2 System, *WSEAS Transactions on Systems*, Issue 11,v. 4, Nov.2005, pp. 1923-1930.

[9] G. Jäger, U. Benz, Measures of Classification Accuracy Based on Fuzzy Similarity, *IEEE Trans. On Geoscience and Remote Sensing,* Vol. 38, No 3, 2000, pp. 1462-1467.

[10] J. Canny, A computational approach to edge detection, *IEEE transactions on Pattern Analysis and Machine Intelligence,* Vol. 8, 1986, pp. 679-698.

[11] A.J. Baddeley, Errors in binary images and an L^p version of the Hausdorff metric, *Nieuw. Arch.Wiskd.,IV,* Ser. 10, No. 3, 1992, pp. 157-183.

[12] M. S. Prieto, A.R. Allen, A Similarity Metrics for Edge Images, *IEEE Trans. on Pattern Analysis and Machine intelligence,* Vol. 25, No 10, 2003, pp. 1265-1273.

[13] R. G. Congalton, A review of assessing the accuracy of classifications of remotely sensed data, *Remote Sensing of Environment,* Vol. 37, No 1, 1991, pp. 35-46.

[14] B.B. Chaudhuri, A. Rosenfeld, A modified Hausdorff distance between fuzzy sets, *Information Sciences,* v. 118, 1999, pp. 159-171.