Fuzzy Hamming Distance Based Banknote Validator

Mircea Ionescu
Department of ECECS,
University of Cincinnati,
Cincinnati, OH 45221-0030, USA
ionescmm@ececs.uc.edu

Anca Ralescu
Department of ECECS,
University of Cincinnati,
Cincinnati, OH 45221-0030, USA
aralescu@ececs.uc.edu

***Abstract*— Banknote validation systems are used to discriminate between genuine and counterfeit banknotes. The paper propose a one-class classifier for genuine class using a new similarity measure, Fuzzy Hamming Distance. For each banknote several regions are considered (corresponding to security features) and each region is split in $m \times n$ partitions, to include position information. The feature space used by the classifier consists from color histograms of each partition. The Fuzzy Hamming Distance proves to have a good discrimination power being able to completely discriminate between the genuine and counterfeit banknotes.**

I. Introduction

Automatic banknote handling is a huge industry including ATMs, banks or vending machines. The key element of automatic banknote handling is the banknote validation system. A banknote validation system check whether the input banknote is genuine or a counterfeit. The input banknote is scanned using different sensors varying from optical scan to magnetic scan. Then the acquired attributes are compared with the templates for each banknote type. If a match is found then the banknote is genuine, otherwise is considered counterfeit.

This problem can be seen as one-class classification because the number of counterfeit examples available is very small. Also it is required that the genuine class is learned very well. This is the approach in Girolami's work [1], where each banknote is divided into $m \times n$ regions. An individual classifier is built for each region, then classifiers are combined to provide an overall decision. To find the best values for m and n a genetic algorithm is used. This approach assures that the system will work with any banknote,i.e. no banknote specific information is used.

Even though this approach simplifies the update of the system, by not using expert knowledge for setting the system, it ignores valuable information regarding security features that are banknote specific.

In the previouse studies, [2], [3], [4], Fuzzy Hamming Distance proved to be efficient in a Content Based Image Retrieval system, that output the closest images in the database given a quey image. The study proposed in this paper investigates further the discrimination power of FHD in the case where the compared objects are very similar, genuie and counterfeit banknote, unlike the CBIR case where the compared objects are just close.

The Fuzzy Hamming Distance based approach, proposed here, combines the versatility of an automatic system with basic banknote specific information. Subsequently, the system can be updated to use in-depth security features provided by an expert.

II. Fuzzy Hamming Distance

A. Introduction

The Fuzzy Hamming Distance is a generalization of Hamming distance over the set of real-valued vectors. It preserves the original meaning of the Hamming distance as the number of different components between the input vectors with the added features that it uses real-valued vectors and it takes in account the amount of the difference between each component. FHD is the (fuzzy) number of different components of the input vectors. The fuzzy set shows the degree to which the input vectors are different by 0, 1,...,n, where n is the size of the vectors. In short, the Fuzzy Hamming Distance is the fuzzy cardinality of the difference fuzzy set. Therefore, in order to define it, one needs to define in order the concepts of *degree of the difference* between two real values, *difference fuzzy set* and make use of the concepts of *cardinality of a fuzzy set* and *defuzzification of a fuzzy set*.

B. Degree of Difference

Definition 1: (**Degree of difference [5]**): Given the real values x and y, the degree of difference between x and y, modulated by $\alpha > 0$, denoted by $d_\alpha(x, y)$, is defined as:

$$d_\alpha(x, y) = 1 - e^{-\alpha(x-y)^2} \quad (1)$$

The parameter $\alpha \geq 0$ *modulates* the degree of difference in the sense that for the same value of $|x - y|$, different values of α will result in different values of $d_\alpha(x, y)$.

C. Difference Fuzzy Set

Using the notion of *degree of difference* defined above, *the difference fuzzy set* $D_\alpha(x, y)$ for two vectors x and y, is defined as:

Definition 2: **(Difference fuzzy set for two vectors [5])**: Let x and y be two n-dimensional real vectors, and let x_i, y_i denote their corresponding i^{th} component. The degree of difference between x and y along the component i, modulated by the parameter α, is $d_\alpha(x_i, y_i)$.

The difference fuzzy set corresponding to $d_\alpha(x_i, y_i)$ is $D_\alpha(x, y)$ with membership function

$\mu_{D_\alpha(x,y)} : \{1, \ldots, n\} \rightarrow [0, 1]$ given by equation 2:

$$\mu_{D_\alpha(x,y)}(i) = d_\alpha(x_i, y_i) \quad (2)$$

In words, $\mu_{D_\alpha(x,y)}(i)$ is the degree to which the vectors x and y are different along their ith component. The properties of $\mu_{D_\alpha(x,y)}$ are determined from the properties of $d_\alpha(x_i, y_i)$ [5].

D. Cardinality of a fuzzy set and Crisp Cardinality

Of subsequent use in this study is the notion of *cardinality of a fuzzy set*. This concept has been studied by several authors starting with [6], and continuing with [7], [8], [9], etc. Here the definition put forward in [8] and further developed in [9] is used.

Definition 3: **(Cardinality of a fuzzy set [8])**:Let

$$A \equiv \sum_{i=1}^{n} x_i/\mu_i$$

denote the discrete fuzzy set A over the universe of discourse $\{x_1, \ldots, x_n\}$ where $\mu_i = \mu_A(x_i)$ denotes the *degree of membership for x_i to A*. The cardinality, $CardA$, of A is a fuzzy set

$$CardA \equiv \sum_{i=0}^{n} i/\mu_{CardA}(i)$$

where

$$\mu_{CardA}(i) = \mu_{(i)} \wedge (1 - \mu_{(i+1)}) \quad (3)$$

where $\mu_{(i)}$ denotes the ith largest value of μ_i, the values $\mu_{(0)} = 1$ and $\mu_{(n+1)} = 0$ are introduced for convenience, and $\wedge$ denotes the min operation.

For the defuzzification of the fuzzy cardinality of a fuzzy set A, the non-fuzzy cardinality $nCard(A)$, introduced in [9] is the only one which guarantees that the result is a whole number, and therefore satisfies its semantics, *"the number of* ...". Starting from a set of requirements on $nCard(A)$, it is shown in [9], that $nCard(A)$ is equal to the cardinality (in classical sense) of $A_{0.5}$ the λ-level set of A with $\lambda = 0.5$. More precisely,

$$nCard(A) \equiv |\overline{\{x; \mu_A(x) > 0.5\}}| \quad (4)$$

where for a set S, $\overline{S}$ denotes the closure of S.

E. The Fuzzy Hamming Distance

By analogy with the definition of the classical Hamming distance, and using the cardinality of a fuzzy set [10], [6], [7], the Fuzzy Hamming Distance (FHD) is now defined as:

Definition 4: **(The Fuzzy Hamming Distance) [5]** Given two n dimensional real-valued vectors, x and y, for which the difference fuzzy set $D_\alpha(x, y)$, with membership function $\mu_{D_\alpha(x,y)}$ defined in (2), the **fuzzy Hamming distance** between x and y, denoted by $FHD_\alpha(x, y)$ is the fuzzy cardinality of the difference fuzzy set, $D_\alpha(x, y)$:

$\mu_{FHD(x,y)}(\,\cdot\,\alpha) : \{0, \ldots, n\} \rightarrow [0, 1]$ denotes the membership function for $FHD_\alpha(x, y)$ corresponding to the parameter α. More precisely,

$$\mu_{FHD(x,y)}(k; \alpha) = \mu_{CardD_\alpha(x,y)}(k) \quad (5)$$

for $k \in \{0, \ldots, n\}$ where $n = |SupportD_\alpha(x, y)|$.

In words, (5) means that for a given value k, $\mu_{FHD(x,y)}(k; \alpha)$ is *the degree to which the vectors x and y are different on exactly k components (with the modulation constant α).*

Example 1: Consider two vectors x and y with $|x - y| = (2, 1, 4, 3, 5)$. Their FHD is the same as that between the vector, 0 and $|x - y| = (2, 1, 4, 3, 5)$ (by Property (4) of $d_\alpha(x, y)$). Therefore, the FHD between x and y is the cardinality of the fuzzy set $D(|x - y|, 0) \equiv 1/0.9817 + 2/0.6321 + 3/1 + 4/0.9999 + 5/1$ obtained according to (3).

For $\alpha = 1$ this is obtained to be $FHD \equiv 0/0 + 1/0 + 2/0.0001 + 3/0.0183 + 4/0.3679 + 5/0.6321$. In words, the meaning of FHD is as follows: the degree to which x and y differ along exactly 0 components (that is, do not differ) is 0, along exactly one component is 0, along exactly two components is 0.00013, along exactly three components is 0.0183, and so on. The difference fuzzy set and the fuzzy Hamming distance for this example are shown in Figures 1 and 2 respectively.

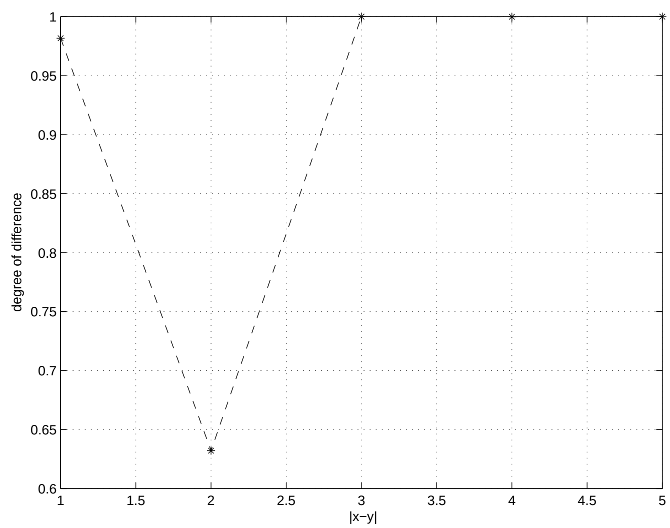

Fig. 1. The difference fuzzy set for the data in Example 1.

The properties of the Fuzzy Hamming Distance follows from its definition as cardinality of a fuzzy set and are summarized in the following proposition:

Proposition 2.1: In the following, x, y, z are real-valued n-dimensional vectors, and $\alpha \in \Re$ and $nFHD$ is the defuzzification of FHD using $nCard$ according to equation 4.

1) $\mu_{FHD(x,y)}(k; \alpha) \geq 0, \forall\ k = 0, \ldots, n + 1;$
2) If $\mu_{FHD(x,y)}(k; \alpha) = 1$ then

$$\mu_{FHD(x,y)}(j; \alpha) = \begin{cases} 1 & \text{if } j = 0, \ldots, k \\ 0 & \text{if } j = k + 1, \ldots, n \end{cases}$$

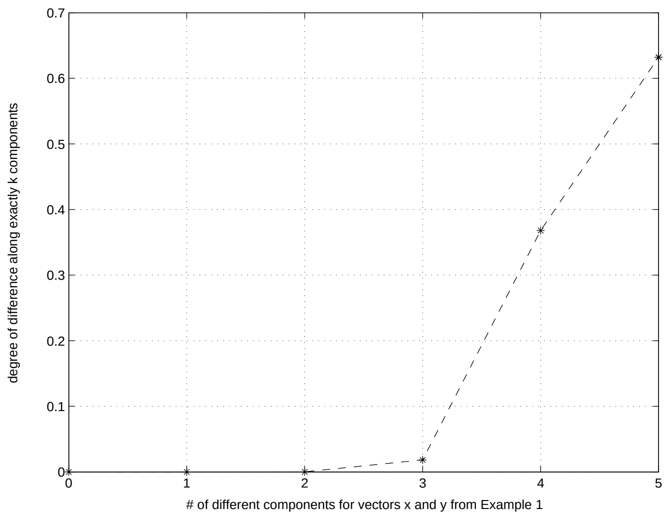

Fig. 2. FHD for the data in Example 1 with the difference fuzzy set shown in Figure 2.

3) $\mu_{FHD(x,x)}(0;\alpha) = 1$ and $\mu_{FHD(x,x)}(k;\alpha) = 0$, $\forall\ k = 1,\ldots,n$;
4) $\mu_{FHD(x,y)}(0;\alpha) = 1 \iff x \equiv y$;
5) $\mu_{FHD(x,y)}(k;\alpha) = \mu_{FHD(y,x)}(k;\alpha)$, $\forall\ k = 0,\ldots,n$;
6) For any $\alpha \in \Re$, there exists at least one permutation x, y, z such that

$$nFHD_\alpha(x,y) \leq nFHD_\alpha(x,z) + nFHD_\alpha(y,z)$$

where $nFHD_\alpha(\cdot,\cdot)$ corresponds $FHD_{\cdot,\cdot}(\cdot,\alpha)$;
7) For any x, y, z there exist $\alpha_1, \alpha_2, \alpha_3$ such that

$$nFHD_{\alpha_1}(x,y) \leq nFHD_{\alpha_2}(x,z) + nFHD_{\alpha_3}(y,z)$$

8) There exists α, such that for binary vectors, the $nCard$ defuzzification of the Fuzzy Hamming Distance coincides with the classical Hamming Distance(HD) between these vectors, that is, $nCardFHD(x,y) = HD(x,y)$

F. Adapting Fuzzy Hamming Distance

The Fuzzy Hamming Distance can be adapted to include a scaling factor and threshold for the amount of the difference necessary for the difference to be considered as a change. Let x_i and y_i are the i^{th} component of the vectors x and y, $x_i, y_i \in [a_i, b_i]$. The length of the domain for x_i and y_i is $L_i = b_i - a_i$

One way to set the value for α is to impose a lower bound on the membership function subject to constraints on the difference between vector components, such as described in (6) and (7).

$$\mu_{D_\alpha(x,y)}(i) \equiv d_\alpha(x_i, y_i) > 1 - \epsilon \tag{6}$$

subject to

$$|x_i - y_i| > M \tag{7}$$

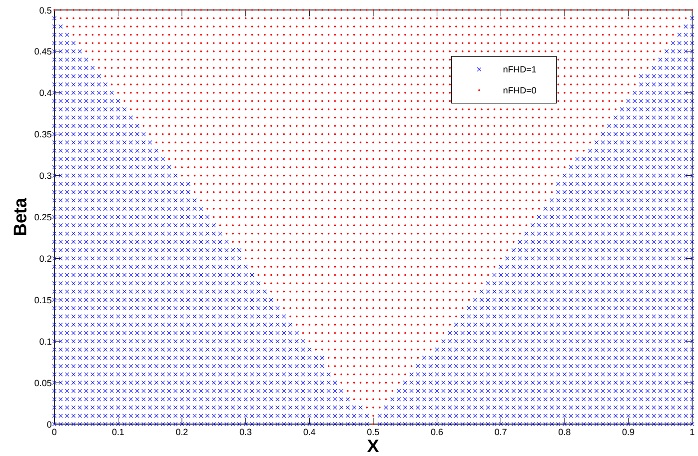

Fig. 3. $nFHD(0.5, x)$ where $x\ in[0, 1]$ for different β

where (x_i, y_i) is a generic component pair, $1 - \epsilon$ a desired lower bound on the membership value to FHD (as defined in (5)), and M some positive constant. In particular, M can be set as $M = \beta L_i$ where $\beta \in [0, 1]$, which leads to

$$1 - e^{-\alpha_i(x_i - y_i)^2} > 1 - \epsilon$$

from which it follows that

$$\alpha_i > \frac{1}{(x_i - y_i)^2} ln\frac{1}{\epsilon} \tag{8}$$

This leads to the following formula for defining the value for the parameter α along the i^{th} component:

$$\alpha_i = \frac{ln\frac{1}{\epsilon}}{L_i^2}\frac{1}{\beta^2} \tag{9}$$

where L_i and β have the meaning stated above. β can be viewed as the percentage from L_i that is considered as a change for that column.

For example, for $\beta = 0.1$ (10%), $L_i = 255$ and $\epsilon = 0.5$, if the difference between compared components of the vectors is greater then 25.5 then the degree of change will be greater than $1 - \epsilon = 0.5$, and therefore it will be counted in defuzzification of FHD.

For $\beta = 0$ if $x_i \neq y_i$ the degree of the difference is 1. Since β can be different for each column, FHD sensitivity can be controlled differently for each column (feature).

Figure 3 shows the nFHD between the point $[0.5]$ and the points in the interval $[0, 1]$ for different $\beta \in [0, 0.5]$.

III. The Dataset

The dataset used is generated by scanning one dollar bills using 600 dpi resolution, as follows:

- For genuine banknotes:
  - 10 scans of the same banknote;
  - one scan of 10 different banknotes;
- For counterfeit banknotes:
  - 10 prints and scans of \$1 banknote;
  - one scan of \$5,\$10,\$20,\$50,\$100 banknote;

Fig. 4. Banknote Partitions

Fig. 5. Normalized Banknote

IV. System Description

The system consists from three modules:

1) Feature Extraction;
2) Training;
3) Classification;

A. Feature Extraction

The feature space used in this study, consists from color histograms but it can be extended to any available space, like the magnetic signature.

The attributes for each banknote are computed in the following steps:

1) *features selection*: split the banknote in seven regions, corresponding to security elements (red squares in Figure 4);
2) *banknote normalization*: replace the areas that are different from one banknote to another, like serial numbers, by black color (orange squares in Figure 4);
3) *partitioning*: each considered region and the normalized banknote is partitioned in $m \times n$ partitions;
4) *color histogram*: compute the 4096 bins color histogram for each partition. The following formula is used to transform the RGB space in a 4096 bins space:

$$Index = \left\lfloor \frac{R}{16} \right\rfloor \times 256 + \left\lfloor \frac{G}{16} \right\rfloor \times 16 + \left\lfloor \frac{B}{16} \right\rfloor$$

In words, the attributes of a banknote consist from the color histogram of the partitions for the seven regions (red regions from Figure 4) plus the histograms for partitions of the entire normalized banknote, Figure 5.

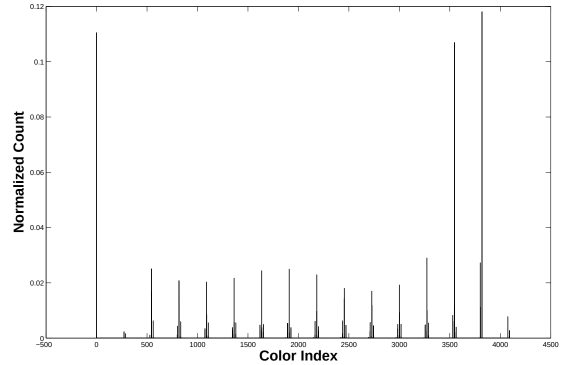

Fig. 6. One of the 32 average color histograms of the normalized banknote, using a 2×2 partitioning

B. Training

The goal of our system is to recognize well one class, the genuine banknotes. Genuine examples are used to train the classifier by computing the the average histograms of each partition on each region. The genuine class prototype (GP) consists from $8 \times m \times n$ average color histograms. Figure 6 shows one of the histograms.

C. Classification

Classification is done based on a threshold classifier using the Fuzzy Hamming Distance. $nFHD$ between the test banknote and the genuine class descriptor is computed and if the distance is less then a certain threshold γ is considered genuine, otherwise is considered counterfeit.

$$D(X, GP) = \frac{1}{N}\sum_{i=1}^{N}\left(\frac{1}{m \cdot n}\sum_{j=1}^{m \cdot n} nFHD(X_{ij}, GP_{ij})\right)$$

where X_{ij} is the color histogram of i^{th} region and j^{th} partition of the test item, N is the number of regions and $nFHD$ is the defuzzification of FHD using crisp cardinality as defined in equation 4.

The classification decision is taken as follows:

$$X = \begin{cases} Genuine & \text{if } D(X, GP) < \gamma \\ Counterfeit & \text{otherwise} \end{cases}$$

V. Results

In order to obtain a robust discrimination of the two classes one needs to find the best β parameter for FHD and also the right partitioning scheme, m and n. Values $\beta = 0, 0.001, 0.002, \ldots, 0.03, 0.04, \ldots 0.4$ and $m \times n$ partitioning for $m = 1, 2, 3, 4$ and $n = 1 \ldots 10$ were used and for all β and $m \times n$ partitioning the distances between the genuine descriptor and the test data, genuine and counterfeit were computed.

For each class, genuine and counterfeit, the minimum, mean, standard deviation, maximum of the distances $D(X, GP)$ are computed and the values β, m, n for which the distance between the genuine class mean and counterfeit class mean was the highest are selected.

$$Margin = Mean_{Genuine} - Mean_{Counterfeit}$$

TABLE I

THE MARGIN IN DECREASING ORDER FOR DIFFERENT VALUES OF β AND $m \times n$ PARTITIONING.

m	n	β	Genuine				Counterfeit				Margin
			Min	**Mean**	σ	**Max**	**Min**	**Mean**	σ	**Max**	
1	1	0.002	16.50	27.12	9.41	51.88	58.88	72.40	4.71	75.63	45.27
1	1	0.001	33.62	43.99	9.58	67.13	77.25	88.21	4.03	91.88	44.22
1	2	0.002	17.94	27.85	9.11	52.50	59.81	71.82	4.29	74.75	43.96
1	2	0.001	34.06	44.50	9.09	67.00	77.00	88.43	4.24	92.26	43.92
2	1	0.002	18.31	27.88	8.92	51.81	60.31	71.17	3.88	74.07	43.29
2	1	0.001	34.44	44.72	9.01	67.00	77.69	87.45	3.67	90.82	42.73
3	1	0.001	34.87	44.92	8.73	67.38	78.00	87.24	3.70	90.76	42.31
3	1	0.002	19.54	28.21	8.51	51.29	60.96	70.51	3.58	73.59	42.29
1	3	0.001	33.92	43.98	8.54	65.34	76.21	85.99	3.71	89.55	42.01
2	2	0.001	35.12	45.30	8.47	67.13	76.85	86.89	3.84	90.69	41.59
1	3	0.002	18.54	27.61	8.22	49.63	58.75	69.20	3.79	72.88	41.59
2	**2**	**0.002**	**19.47**	**28.69**	**8.41**	**51.47**	**59.31**	**70.07**	**3.90**	**72.94**	**41.38**
1	1	0.003	9.876	19.07	8.40	42.13	48.13	60.37	4.34	64.25	41.29
4	1	0.001	35.28	45.28	8.38	66.60	78.00	86.35	3.46	89.88	41.07
1	4	0.001	35.37	44.51	8.00	64.44	76.00	85.41	3.69	88.91	40.89

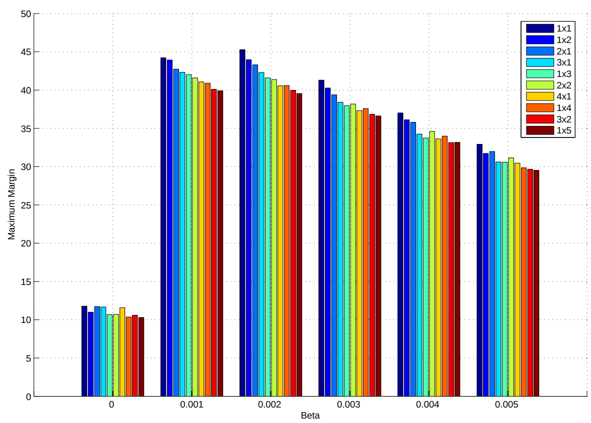

Fig. 7. Top 10 margin for different β and $m \times n$ partitioning

In words, the $Margin$ shows how far apart the genuine and counterfeit classes are, if each class is described by the mean. It is used to find the minimum threshold for the distance to genuine class descriptor to completely recognize the genuine class.

The results are shown in Table I and Figure 7.

To avoid overfitting, the highest values for β, m, n that yield an acceptable margin are selected: $\beta = 0.002$ and $m = n = 2$. For these values the upper bound for the distance between the test data and the genuine class descriptor is $\gamma = 58$. In this case the probability for a genuine banknote to be classified as genuine is $P(0 < x < 58) = 0.999431$, considering the normal distribution with mean=28.69 and $\sigma = 8.41$.

It can be seen from the Table I that a good separation between genuine and counterfeit class for all 15 combination of β and partitioning is obtained. The minimum value of counterfeit class is always greater than the maximum value of the genuine class. Also for a threshold γ, between those two values the probability for a genuine class to be classified as genuine equals 1.

Also from the ROC curve in Figure 8, it follows that a complete discrimination for $\gamma \in [52, 60]$ is obtained; higher values for m and n are more desirable because this way more

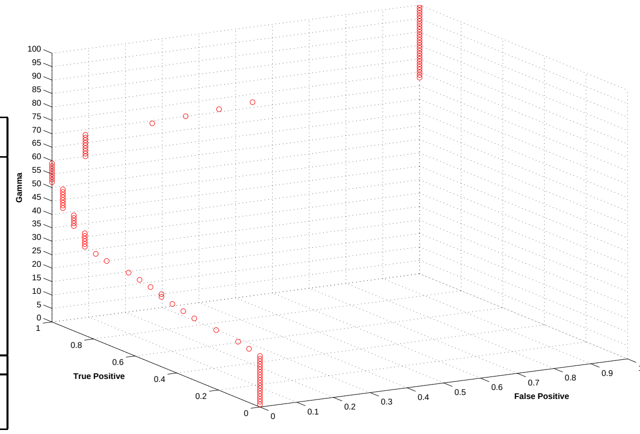

Fig. 8. ROC curve for $\gamma = 1 \ldots 100$, $\beta = 0.002$ and 2×2 partitioning

position information (not only color information) is included.

VI. CONCLUSIONS

The system performed well being able to completely discriminate between the genuine and counterfeit examples. Also is easy to update when new banknotes are introduced in circulation, more genuine example are provided or new security features (feature space) are included in the system.

It can be concluded that FHD proved to have a good selectivity power, being able to distinguish between very similar objects. Changing the values of the parameter β leads to adapting the sensitivity of FHD to the specifics of the application.

Further studies would be needed to extend the training data set using more examples as well as other features such as magnetic image. Also in this study only the defuzzification of FHD is used in order to compute the dissimilarity between the test data and the genuine prototype. Further studies should use FHD directly as a fuzzy set without any defuzzification.

REFERENCES

[1] C. He, M. Girolami, and G. Ross, “Employing optimized combinations of one-class classifiers for automated currency validation,” *Pattern Recognition*, vol. 37, no. 6, pp. 1085–1096, June 2004.

[2] M. M. Ionescu and A. L. Ralescu, “The impact of image partition granularity using fuzzy hamming distance as image similarity measure.” in *MAICS*, April 2004, pp. 111–118.

[3] ——, “Fuzzy hamming distance in a content-based image retrieval system.” in *FUZZ-IEEE*, July 2004.

[4] ——, “Fuzzy hamming distance as image similarity measure.” in *IPMU*, July 2004, pp. 1517–1523.

[5] A. Ralescu, “Generalization of the hamming distance using fuzzy sets,” Research Report, JSPS Senior Research Fellowship, Laboratory for Mathematical Neuroscience, The Brain Science Institute, RIKEN, Japan, May-June 2003.

[6] N. Blanchard, “Cardinal and ordinal theories about fuzzy sets,” Fuzzy Information Processing and Decision Processes,149-157, pp. 149–157, 1982.

[7] L. Zadeh, “A computational approach to fuzzy quantifiers in natural languages,” *Comput. Math.*, vol. 9, pp. 149–184, 1983.

[8] A. Ralescu, “A note on rule representation on expert systems,” *Information Science*, vol. 38, pp. 193–203, 1986.

[9] D. Ralescu, “Cardinality, quantifiers, and the aggregation of fuzzy criteria,” *Fuzzy Sets and Systems*, vol. 69, pp. 355–365, 1995.

[10] R. Hamming, “Error detecting and error correcting codes,” *The Bell System Technical Journal*, vol. XXVI, pp. 147–160, April 1950.