مجلة جامعة تشرين للبحوث والدراسات العلمية _ سلسلة العلوم الهندسية المجلد (32) العدد (6) 2010
Tishreen University Journal for Research and Scientific Studies - Engineering Sciences Series Vol. (32) No. (6) 2010

Efficiency of Using Multi-Sets of Features Technique with Large Number of Features in Automatic Signature Verification

Dr. Maan Ammar*

(Received 18 / 4 / 2010. Accepted 13 / 6 / 2010)

□ ABSTRACT □

This paper introduces the experimental results of using multi-sets of features technique for automatic signature verification with large number of features. The experimental results are analyzed and discussed. The analysis of the results have shown that the multi-sets of features technique remains effective even if the number of used features is large, and the overall performance improves in comparison with smaller number of features. The effect of verification using the best feature set, as well as multi-sets of features is also explored. The reached result is further improvement in the performance.

Key words: Signature verification, feature selection, best feature set, multi-feature set.

***Professor , Biomedical Engineering Department, Faculty of Mechanical and Electrical Engineering, Damascus University.**

مجلة جامعة تشرين للبحوث والدراسات العلمية _ سلسلة العلوم الهندسية المجلد (32) العدد (6) 2010
Tishreen University Journal for Research and Scientific Studies - Engineering Sciences Series Vol. (32) No. (6) 2010

فعالية استخدام تقنية مجموعات الخصائص المتعددة
مع عدد كبير من الخصائص في التحقق الآلي من صحة التواقيع

الدكتور معن عمار*

(تاريخ الإيداع 18 / 4 / 2010. قُبِل للنشر في 13/ 6 / 2010)

□ ملخّص □

تقدم ورقة البحث هذه النتائج التجريبية لاستخدام تقنية مجموعات الخصائص المتعددة الجديدة من أجل التحقق الآلي من صحة التواقيع باستخدام عدد كبير من الخصائص. و يتم في هذه الورقة تحليل النتائج العملية ومناقشتها. لقد بين تحليل النتائج التجريبية أن تقنية مجموعات الخصائص المتعددة تبقى فعالة حتى عندما يكون عدد الخصائص المستخدم كبيرا، وأن الأداء الكلي يتحسن بالمقارنة مع حال استخدام عدد أقل من الخصائص. كما يتم في هذه الورقة استكشاف أثر استخدام أفضل مجموعة خصائص إضافة إلى مجموعات الخصائص المتعددة على نتيجة التحقق من صحة التواقيع. وقد بينت النتائج العملية أن الأثر هو تحسين إضافي للأداء في التحقق الآلي من صحة التواقيع.

الكلمات المفتاحية: التحقق من صحة التواقيع، انتقاء الخصائص، أفضل مجموعة خصائص، مجموعات الخصائص المتعددة.

Introduction:

*أستاذ – قسم الهندسة الطبية، كلية الهندسة الميكانيكية والكهربائية، جامعة دمشق–دمشق–سورية.

Signatures are two types: (1) on-line signatures like those obtained from three axis writing pen (x, y, and pressure), and (2) off-line ones like those we usually find on letters, contracts, and bank checks. The second type is the one dealt with in this paper.

Due to the importance of signatures, they are the target of forgers. Forged signatures are mainly 3 types: (1) **random forgery** in which a completely different signature is used instead of the signature of the specific person. This type of forgeries can be simply detected by a human or by the computer with a suitable program; (2) **simple forgery** in which the forger tries to simulate the signature of the specific person with some effort. This kind of forgeries resembles the genuine signature to some extent, but can easily be detected by an expert and computer. Specialized computer programs can realize high performance in detection of such forgeries; and (3) **skilled forgery** in which the forger practice the signing process until he becomes convinced that he can create a genuine-like forgery. This third kind of forgeries is difficult to be detected by a human or by the computer. Fig. 1 shows examples of simple and skilled forgeries. Verifying signatures by computer (answering the question: is the signature "Genuine" or a "Forgery") is usually called Automatic Signature Verification (ASV).

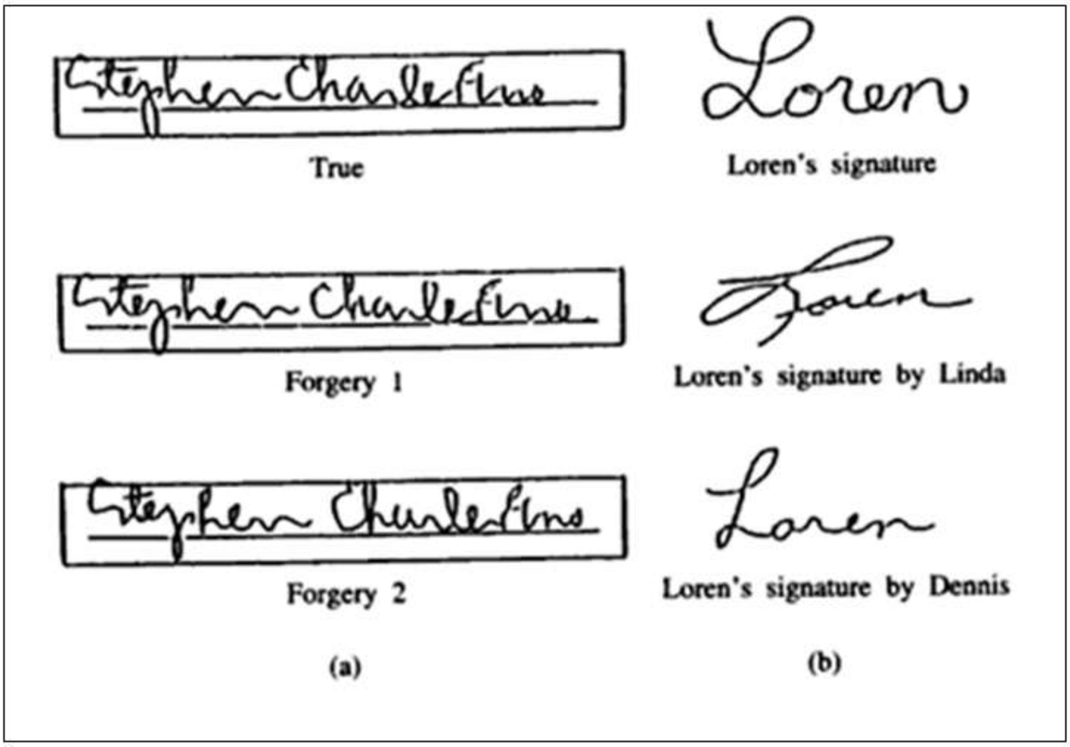

Fig. (1) The upper two signatures are genuine. The two signatures below the upper left one are skilled forgeries, and the two signatures below the right one are simple forgeries [8].

(AMMAR et al.,1986) reported the first successful work on verification of skilled forgeries[1,2]. Their principle of extracting High Pressure Regions (HPRs) in signatures was adopted later by other researchers for further study [3,4]. It has also motivated others to explore other ways of determining the threshold used to extract the HPRs [5,6]. In 1989 they investigated using shape features, HPR features, and both of them for ASV [7]. At the same time, they investigated the effectiveness of individual shape features, different shape feature sets, and mixed ones with the implications of automatic determination of the verification threshold VTH, using a feature selection algorithm they developed [8]. AMMAR used signature projections and matching for extracting new features [9]. He investigated the performance of the new features and all previous ones using the same feature selection algorithm explained in [8] and reached new results [9,10]. Later, other researchers used projections and shape features introduced previously by AMMAR et al. [8] like baseline, area, and the ratio of height and width for off-line ASV with different decision making approach [11,12]. Recently, some research works attempt to practically evaluate published approaches[13], and others are reattempting to explore the potential effectiveness in the gray level image [14]. This research was done at Faculty of Information Engineering, Nagoya university, Nagoya, Japan in the period 27/11/2009 to 28/2/2010. It can be considered as a continuation of the previous research works related to signatures done at the same faculty.

Aim And Objectives:

Since handwritten signatures are used to authorize important and valuable documents like contracts, and bank checks, they are the target of forgers. In the U.S.A. alone, financial institutions lose 12 billion Dollars due to forgery documents according to American Bankers Association, 1998. At the same time, Americans write about 60 billion checks a year. Therefore, automating forged signature detection by computer is an important requirement.

AMMAR have shown that Multi-Sets of Features (MSF) decision making technique can provide important improvement in detection of skilled forgeries by using rather small number of features (12 features) for feature selection and distance measure [15].

For further improvement of the performance of ASV systems, this paper, investigates the efficiency of using the MSF decision making technique **with large number of features**, and evaluates the effectiveness of using the best feature set (bfs) as well as the MSF in ASV and reaches important results.

Materials And Methods:

The materials used in this research are the signatures available in the signature database, and the methods are the algorithms used to realize the objectives of the research.

1. Signature data

The signature data used in this research consists of 560 genuine and forgery signatures belong to 26 writers. The signatures are written in different languages by people of different nationalities including Arabic, Japanese, Koreans, Europeans, and Americans. Fig.(2) shows examples of the signatures in the database. The number of genuine signatures and forgeries differ from one person to another. Moreover, the documents from which the signatures were extracted vary from white paper, business documents, to bank checks so that the signature data is naturally written under widely different conditions. Forgeries were created with a good attention in order to have convincing forgeries, and some forgeries are real ones obtained from actual caseworks. Fig. (3) shows a group of 10 forgeries, and 6 genuine signatures (last 6 samples in the Figure) of the same person of the used signature data. It is clear that the forgeries are skilled to a good degree.

2. Feature extraction

The features used in this paper are a modified version of the previous ones reported in [8]. Specifically, they are the four slants (positive, negative, vertical and horizontal) measured locally on the contour-detected-signature divided horizontally into six parts, and globally on the image as a whole. The six parts are determined as 3 equal length parts to the left of the Gravity Center of the signature, and 3 equal width others to its right; the "x" and "y" coordinates of the Gravity Center; effective length: the length containing 80% of the area of the signature after omitting 10% to the left and 10% to the right; effective width computed in a way similar to the length; the baseline; and the area of the signature in each one of the six parts computed as a percentage of the total area.

Fig. (2) Examples of the signatures available in the signature database.

Fig. (3) Examples from the signature data used in the experiments. The first 10 signatures are forgeries, and the remaining (last) 6 signatures are genuine.

3. Distance measure and verification decision

The Distance Measure (DM) measures the similarity between the input signature and the reference one(s). The Euclidean distance is used for this purpose in this research. It is computed from the features using eq. (1).

$$DM = (1/n \sum_{i=1}^{n} (f_i-\mu_i/\sigma_i)^2)^{1/2} \qquad (1)$$

Where:

f_i: the i^{th} feature($1 \leq i \leq n$).

n: number of used features.

μ_i: the mean the i^{th} feature computed on the set of genuine (training) samples of the related person.

σ_i: the standard deviation of the i^{th} feature computed on the same set.

The verification decision is made as follows:

If DM > VTH, the input signature is judged to be "genuine", otherwise, it is judged to be "an attempted forgery". VTH is the Verification Threshold.

The value of the VTH is usually determined based on some evaluation experiments using a reference signature data, like that explained in section 2, so that it minimizes the error rate (maximizes the correct decisions).

Determining the used features is usually done either based on the developer experience (not very accurate, but works), or based on a feature selection technique that selects the best feature set (bfs). bfs is the feature set that gives the highest performance. AMMAR et al. developed a feature selection technique based on the principle of the "Circulant Matrix" (Circulant Matrix-Based Feature Selection Technique CMBFST) to generate n^2 feature sets among the possible n! feature sets of n given features, and found that evaluating the signature data available using these n^2 feature sets will lead to the best one after, at most, one or two shuffling processes of the initial order of the features $\{f_1, f_2, \ldots, f_n\}$ [8]. This CMBFST is a very fast one and gives a clear idea of the effectiveness of the individual features, and their contribution to the effectiveness of the different feature sets if augmented by to form a new one.

In order to evaluate the ASV experimental results, we need to define three quantities: PCA, PCR and SR where:

PCA: Percentage of Correct Acceptance (percentage of genuine signatures accepted as genuine samples).

PCR: Percentage of Correct Rejection (percentage of forgeries rejected and classified as attempted forgeries).

SR: System Reliability = (PCA+PCR)/2.

AMMAR have shown that the new MSF technique gives important improvement in the performance of ASV systems using 12 features in the CMBFST-based evaluation . Fig. (4) shows the SR, PCA and PCR curves of the bfs obtained from 12 features using the CMBFST [15].

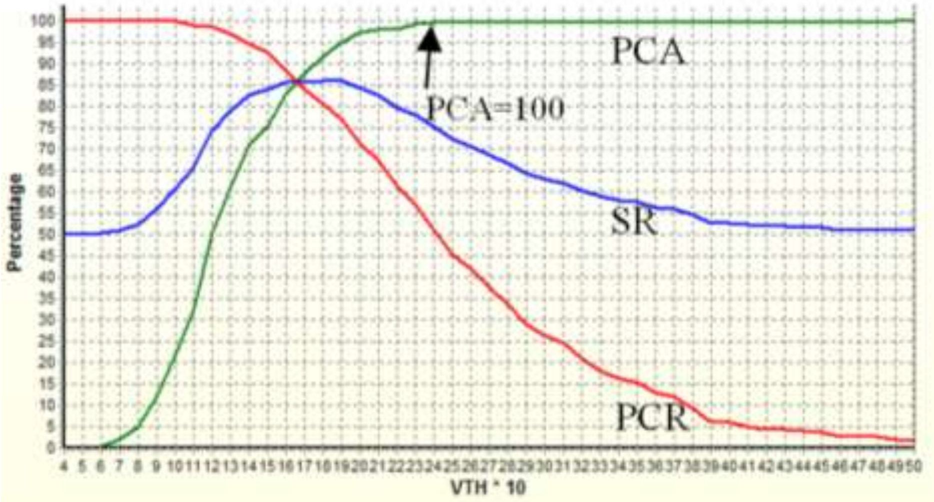

Fig.(4) SR, PCA and PCR curves of the bfs obtained from 12 features[15].

4. The MSF Technique

The new MSF technique reported in detail in [15] depends on verification using "m" feature sets and gathering the detected forgeries. The "m" feature sets are those close in performance to the bfs (called Effective Feature Sets EFS). They are found using the CMBFST. This technique is summarized in Fig. (5) shown below.

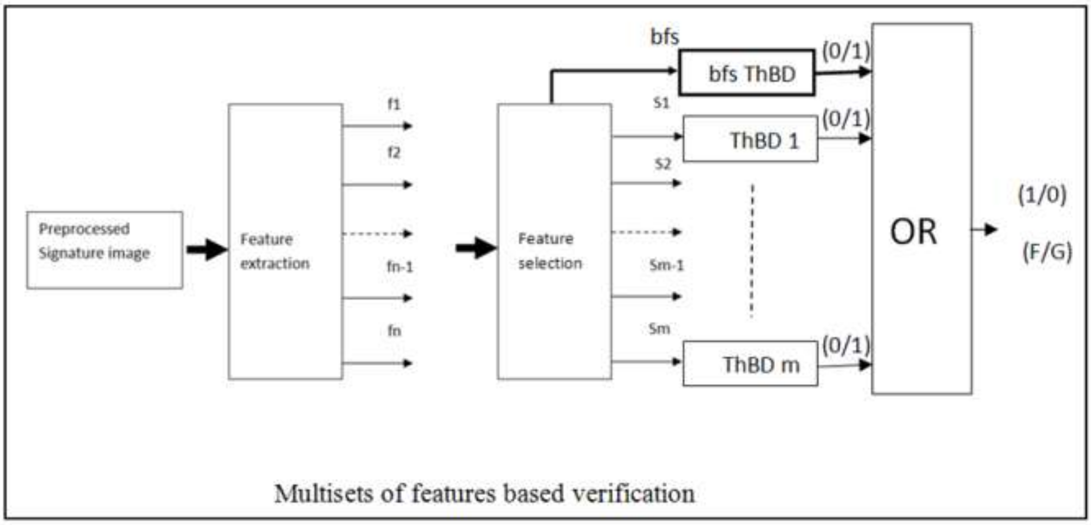

Fig. (5) the MSF technique. Symbols are:{f: feature; S1,..,Sm: EFS; ThBD: Threshold-Based Decision; F: Forgery; G: Genuine}.

This process of verification using the MSF will lead to introducing some error with every feature set used if the VTH goes down below PCA=100 limit: (VTH = 2.4) in Fig. 4, for example. When VTH goes lower than that limit, we will loose in PCA, but will gain in PCR so that the total effect will be positive and in favor of PCR until some VTH value (VTH=1.85 in Fig. 6). Fig. (6) shows the performance of MSF using 18 EFS, and bfs obtained from initial 12 features. The thick curves are for the MSF, and thin ones are for the bfs. The advantage of MSF over bfs is clear where SR of MSF is higher for VTH>1.85. This Figure will be used for comparison with the performance of the MSF in case of large number of features used to find the EFS by the CMBFST.

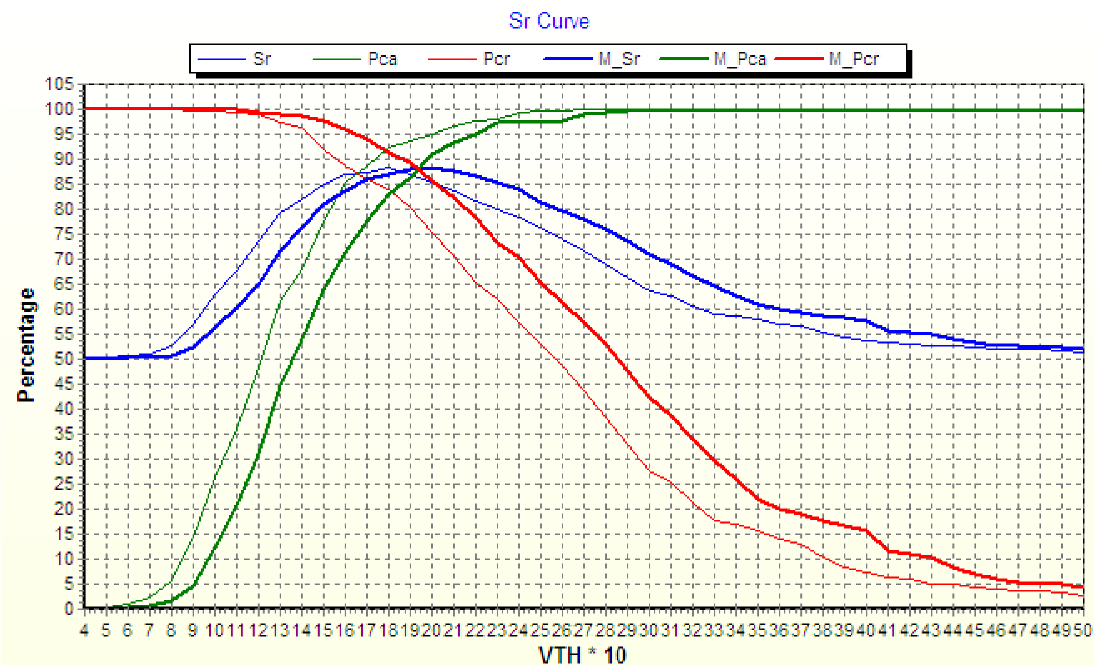

Fig. (6) SR, PCA, and PCR curves of the bfs and MSF obtained from 12 features.

Results And Discussion:

1. Experimental results of using large number of features

In this paper, the performance of the MSF is studied using a large number of features (65 features). Fig. (7) shows a part of the result matrix of 625 entries and 65 features. The symbols appearing in the result matrix denoting the features are explained as follows:

a6, p6, n6, v6, h6 are: area and percentage of positively, negatively, vertically, and horizontally slanted pixels in the 6 parts of the signature explained in section 2.1.

a3, p3, n3, v3, h3 are: the same features above but computed on the signature thirds resulting from combining each two sixths starting from left to the right.

A2, p2, n2, v2, h2 are: the same features above but computed on the signature halve to the left and right of the Gravity Center.

It is worth noting here that "a6" means the 6 local areas measured on the six parts of the signature(6 features), and "p6" means the percentage of the positively slanted pixels in the six parts of the signature (6 features). Since no part of the signature can be omitted, the result matrix considered the local features as groups (6 local areas, 6 percentage of positively slanted pixels, 3 local areas, and so on.) Therefore, the total number of features actually used in the result matrix at the last column is 65 (the last 19 columns appear in the screen shot in Fig. 7. Since the best feature set is chosen to maximize SR, it will be that of the entry (1,20) which gives SR=90.9.

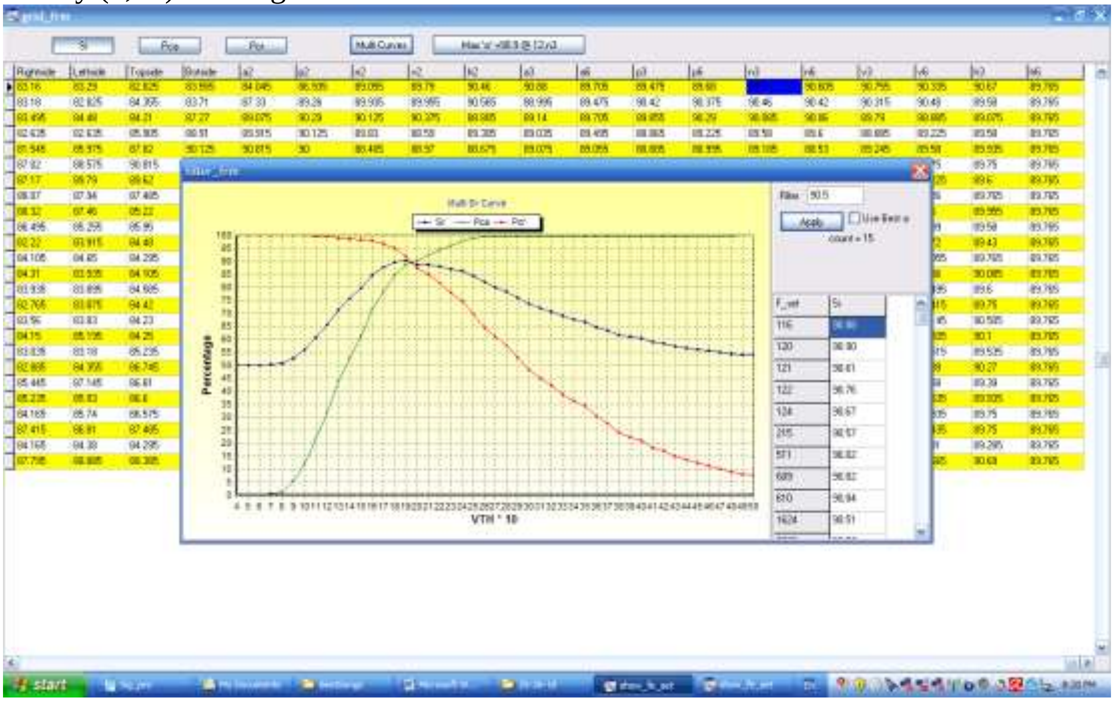

Fig. (7) A part of the SR result matrix of (25x25) entries and 65 features, and the curves of the MSF of 15 EFS with SR>90.5.

2. Performance of the MSF with 625 entries SR matrix

In order to investigate the behavior of the MSF technique with larger number of features, it was tested with 625, (25x25), entries result matrices and 65 features. A part of the SR matrix is shown in the screen shot with the curves of the obtained bfs in Fig. (7). The bfs obtained is marked at (1,20) entry with SR=90.9. The MSF curves corresponding to this bfs and obtained by using 15 EFS with SR>90.5 as appears in the screen shot shown in Fig. (7), are shown in n Fig. (8) with those of bfs. In these curves, we find that:

1 – the performance of the bfs obtained from the 625 entries result matrix and 65 features is considerably better than that obtained from 144 entries and 12 features in Fig. (6), (90.9 in comparison with 88.17).

2 – The MSF performance is also considerably better than that obtained from the 144 entries and 12 features (90.31, in comparison with 88.26), as appears in Fig. (9).

3 – In Fig. (8), although the peak of MSF curve is little bit lower than that of the bfs (90.31 in comparison with 90.9), the peak of the MFS is more flat. Consequently, it is more convenient for selection of the VTH for practical use.

4 – The PCA curve of the MSF is more smooth and approached better the PCA curve of the bfs. Consequently, it is more convenient for making a "Zero false alarm" decision. In Fig. (8), we can get a "Zero false alarm" decision at PCR=70, in comparison with PCR= 50 in Fig. (6). This result leads us to the fact that: **"increasing the number of used features in the primary feature set, and with the proper selection, we can get better performance".**

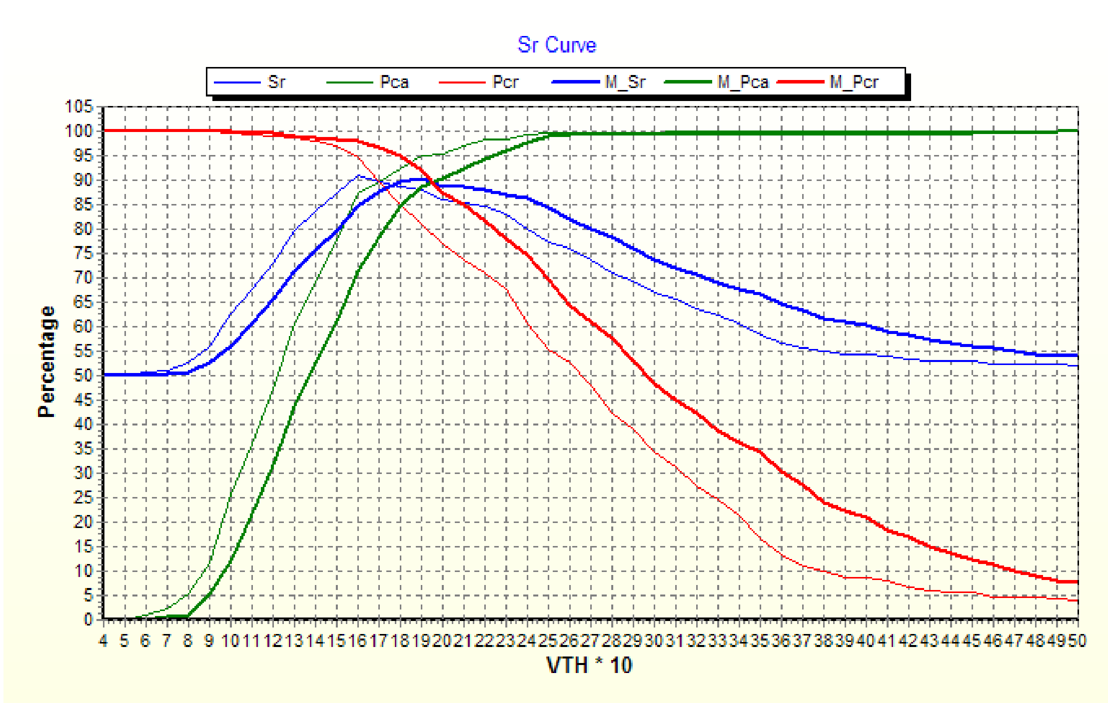

Fig. (8) SR, PCA and PCR curves of bfs (thin ones), and MSF of 15 EFS (thick ones).

5 – The **pure gain in forgery detection** remained the same: about 15% on the PCR scale).

6 – In general, and as the curves in Fig. (8) in comparison with Fig. (6) show, the gained performance in forgery detection with the MSF is better. **This finding reflects the fact that the MSF technique with feature set containing larger number of properly selected features, provide better forgery detection.**

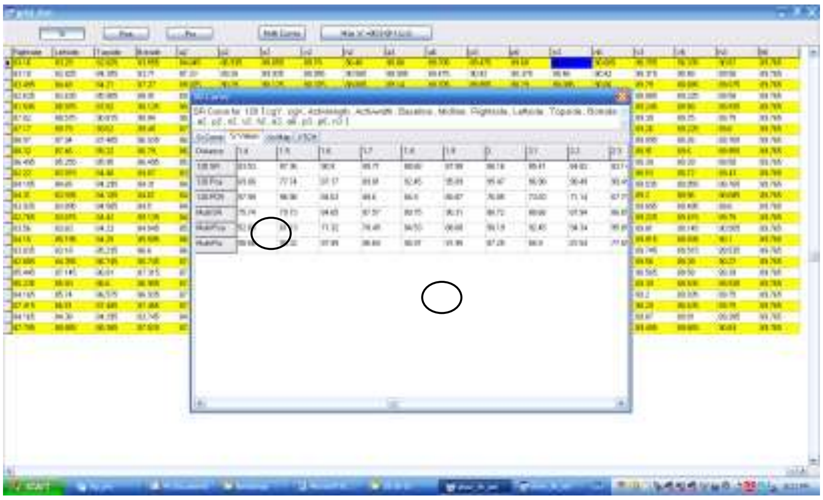

Fig. (9) the values of PCA, PCR and SR of MSF and bfs near the peaks of SR.

3. Performance of the MSF with the bfs

In order to evaluate the efficiency of using the MSF with the bfs, the verification procedure of Fig.(5) (including the bold block of bfs) was tested. The result came interesting as shown in Fig. (10) where:

1 – We got higher SR_{max} = 91.28 (shown in Fig. 11) in comparison with 90.9 for the bfs, and 90.31 for the MSF.

2 - This higher SR is obtained at distinctly high PCR approaching 95%.

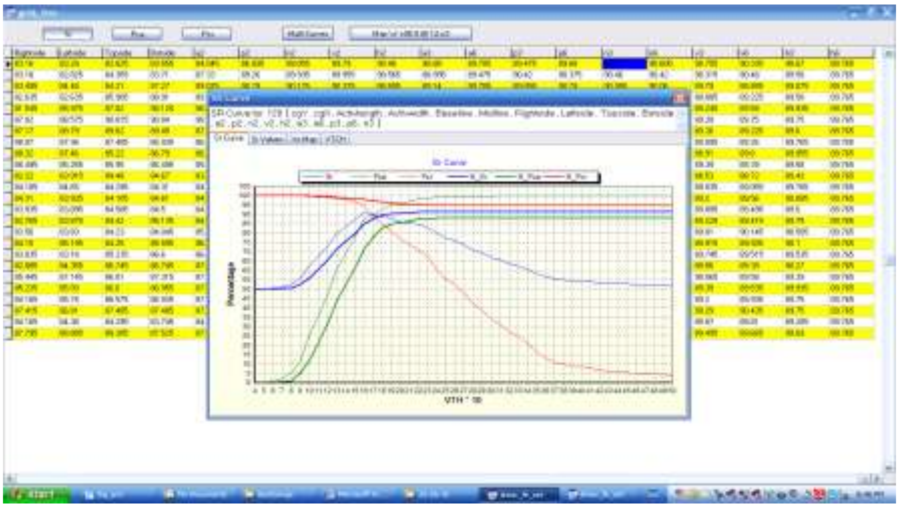

Fig. (10) The PCA, PCR and SR curves of verification with bfs and (bfs+MSF).

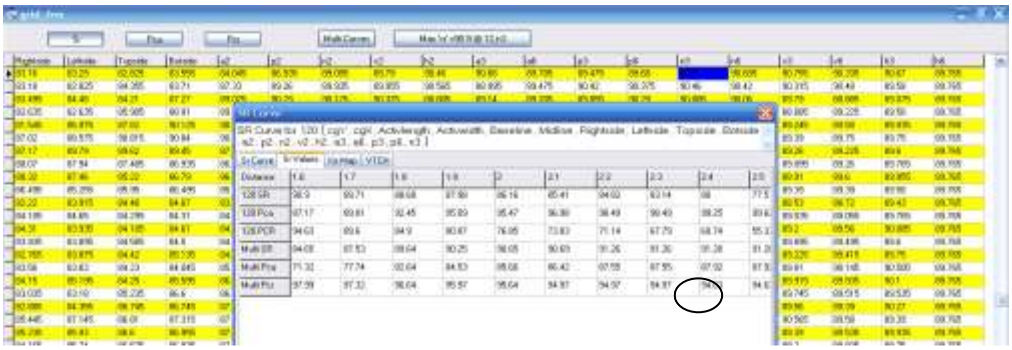

Fig. (11) the values of PCA, PCR and SR of bfs and (MSF+bfs) near the peaks of SR.

4. Summary of the performance improvement with MSF

We can summarize the performance improvements that could obtained by using the MSF technique as follows:

1 – Over 25% relative improvement in forgery detection (15% in PCA) in comparison with the best feature set performance over a wide range of VTH.

2- More flat peak of the SR curve.

3 – A zero false alarm has been possible at a considerably higher PCR.

4 – The improvement obtained by the MSF is very important especially if we keep in mind that :

A – it is **over** the performance of the **best feature set**.

B – any improvement could be important if we are dealing with high value documents or checks like that in Fig. 12 ($35000 forgery check) detected by a software using ASV[15,16].

C – the data used in the experiments are collected from real documents, and the forgeries are skilled to a good extent, as Fig. 2 shows.

D – This improvement obtained by the MSF is **usually lost** in the common single feature set approach.

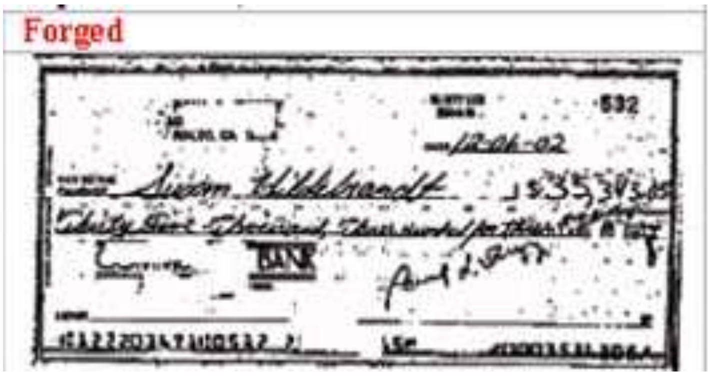

Fig. (12) A thirty five American Dollars check detected automatically by an ASV software.

CONCLUSIONS:

This paper has introduced the results of using the new MSF technique in signature verification using large number of features. Analysis of the experimental results have shown that the MSF remains effective even if the number of used features increase. It was also found that the overall performance improves. Exploring the effect of using the bfs with the MSF in verification, increases the performance further. The importance of this improvement obtained by using the MSF is that it is usually lost in common single feature set based verification. The MSF is a general approach, and not necessarily restricted to signatures.

References:

1. AMMAR M., YOSHIDA Y. and FKUMURA T., *Automatic Off-line Verification of Signatures Based on Pressure Features*, IEEE, Trans on Systems Man and Cybernetics, U.S.A. Vol. SMC-16, No 3, 1986, 39-47.
2. AMMAR M., YOSHIDA Y. and FUKUMURA T., *A New Effective Approach for Automatic Off-line Verification of Signatures by Using Pressure Features*, Proceedings of the 8th Int. Conf. on Pattern Recognition, Paris, France, 1986, 566-569.
3. HUANG K. and HONG Y., *Off-line signature verification based on geometric feature extraction and neural network classification*, Patten Recognition, U.K. Vol. 30, No. 1, 1997, 9-17.
4. SANSONE C. and VENTO M., *Signature verification: increasing performance by a multi-stage system*, Pattern Analysis and Applications, Vol. 3, 2000, 169-181.
5. FRANCISCO J. et al., *Off-line signature verification based on high pressure polar distribution*, 11th ICFHR, 2008, 373-378.
6. MITRA A. et al., *Automatic Authentication of handwritten documents via low density pixel measurements*, International Journal of Computational Intelligence, Vol. 2, No. 4, 2005, 219-223.
7. AMMAR M. et al., *Off-line preprocessing and verification of signatures*, International Journal of Pattern Recognition and Artificial Intelligence, USA, Vol. 2, 1989, 589-602.
8. AMMAR M. et al., *Feature extraction and selection for simulated signature verification*, Computer Recognition and Human Production of Handwriting", R. Plamondon et al. (editors), World scientific Publishing, Singapore, 1989, 61-76.
9. AMMAR M., *Performance of Parametric and Reference Pattern Based Features in Static Signature Verification: A Comparative Study*, Proc. of the 10th ICPR, Atlantic City, U.S.A., 1990, 646-649.
10. AMMAR M., *Progress in verification of skillfully simulated handwritten signatures*, IJPRAI, Singapore, USA , Vol. 5, No. 1&2, 1991, 337-351.
11. FRIAS E. et al., *Support vector machines versus multi-layer perceptrons for efficient off-line signature verification*, Engineering Applications of Artificial Intelligence, Italy, Vol. 19, 2006, 693-704.
12. NGUYEN V. et al., *Off-line Signature Verification Using Enhanced Modified Direction Features in Conjunction with Neural Classifiers and Support Vector Machines*, ICDAR, 2007, 734-738.
13. LARKIN L., *Off-line signature verification*, Doctor thesis, University of Waikato, New Zealand, 2009, 185.
14. JESUS F., et al., *Off-line signature verification based on pseudo-capstral coefficients*, 10th ICDAR, Spain, 2009, 126-130.
15. AMMAR M., et al., A New decision making approach for improving the performance of automatic signature verification using multi-sets of features, Accepted, International Conference of Frontiers on Handwriting Recognition, 2010 (ICFHR 2010), kalekotta, India, October, 2010.
16. AMMAR M., *Method and apparatus for verification of signatures*, United States Patent: No. 6424728 , 07/23/2002.