Transitivity Based Enrollment Strategy for Signature Verification Systems

Claus Vielhauer[1,2], Ralf Steinmetz[1,3] Astrid Mayerhöfer[2,3]
1 – Technical University Darmstadt – Institute for Industrial Process- and System Communications (KOM)
Merckstrasse 25, D-64283 Darmstadt, Germany
2 – Platanista GmbH
Pankratiusstrasse 7, D-64289 Darmstadt, Germany
3 – GMD IPSI
Dolivostrasse 15, D-64293 Darmstadt, Germany
{Claus.Vielhauer, Ralf.Steinmetz}@KOM.tu-darmstadt.de, mayerh@darmstadt.gmd.de

Abstract

The enrollment phase of signature verification systems is a critical process, in which reference data of a user is acquired, that needs to be of satisfying quality without overloading the subject by asking for too many repetitions. Many signature verification systems do not perform an enrollment quality evaluation at all, or only after capturing a fixed number of samples, accepting or rejecting the whole reference set. To limit the number of rejections and as such the False-Enrollment-Rate (FER), we propose a new algorithm for sample adaptive quality evaluation during the enrollment process. This algorithm is based on transitivity criteria within a set of multi-dimensional reference vectors. We will show that our approach leads to a significant reduction of FER.

1. Introduction

Biometric verification systems based on human handwriting have been a research subject since many years and a wide variety of approaches for user verification based on online signature analysis exist, for example [1,2,3]. Typically, error rates of such systems are referred to as a quality measurement, where particularly the error classes of False-Rejection-Rates (FRR) and False-Acceptance-Rates (FAR) are in the focus of interest. However, another critical process of all signature verification systems is the enrollment, which feeds a verification system with reference data for each user. The difficulty is to find criteria to decide if the reference data captured during enrollment are of satisfying quality and at the same time not to overload the user by requesting too many signature samples. Tests have shown that asking too many signature samples from an user can lead to discouraged users and thus to a great variance between the signature samples within a single enrollment. Appropriate enrollment strategies are required to solve this trade-off problem.

Many signature verification systems do not address the question of enrollment quality at all and built references based on a fixed number of handwriting samples during enrollment. Quality measurement of the reference data is performed only after completion of the last sample [4, 5]. If the enrollment does not fulfill the quality criteria, the user is asked to perform the whole enrollment process all over again. This strategy shows the disadvantage that the user has to repeat the whole process again, although possibly only one sample may generate an unacceptable variance.

Another approach described in [6] is based on classifiers that are obtained from only one reference signature sample. It selects one particular reference signature from a set of signatures collected during enrollment. However this approach does not give an indication as how many samples in the set would lead to an optimal solution.

Measurement for the overall enrollment performance of biometric systems is the False-Enrollment-Error (FER) that is the ratio of unsuccessful enrollment attempts towards the total number of enrollments. In order to limit the FER and the total number of user signatures required for successful construction of a reliable reference data set, adaptive enrollment procedures are required, that detect weak handwriting samples within a set of several and allow refinement of reference quality by a well-aimed replacement.

We propose a new algorithm for enrollment evaluation based on transitivity within a set of multi-dimensional reference vectors and will show in our first test results, that application of this algorithm already shows significant improvements of FER. Further, we have indications, that this approach, although originally designed only with the aim to optimize FER, may also improve FAR/FRR rates.

Proceedings of the Sixth International Conference on Document Analysis and Recognition (ICDAR'01)
0-7695-1263-1/01 $10.00 © 2001 IEEE

2. System Overview

Our signature verification system is an online-based system that acquires the physical input signal from a digitizer tablet. Input signals processed by the acquisition module are the pen position signals x(t), y(t) and the binary pressure signal p(t) (pen-down/pen-up characteristics).

Feature extraction modules calculate characteristics from the input signals, which are represented by feature vectors, as shown in the following figure.

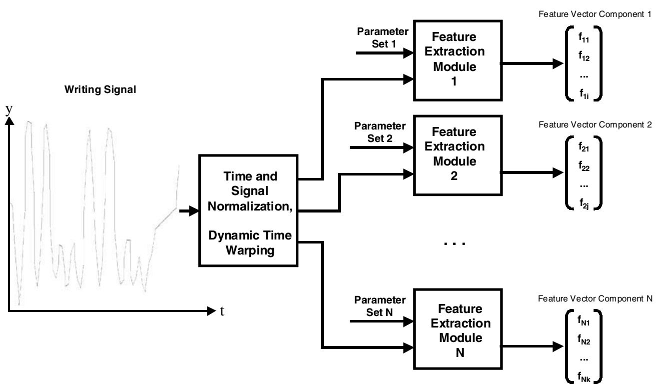

Figure 1 Feature extraction based on writing signal y(t)

Although a more extensive range of writing signals like velocity, acceleration or quantified pressure signals can be considered for user verification, we are currently limiting the signals which are considered for the decision process. In our first test scenario, we are only looking at the pen position signals x(t) and y(t).

The enrollment process uses multiple feature vectors to collect a reference data set. This reference data set is input to a reference evaluation module, which decides whether the reference will be accepted, rejected (leading to an enrollment error) or additional reference samples will be requested from the user. The following chart outlines the enrollment process.

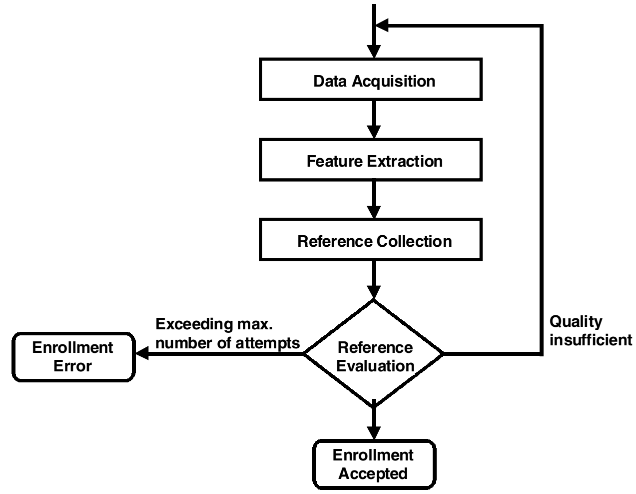

Figure 2 Enrollment process

3 Enrollment Algorithm based on the Strategy of Distance Value Transitivity

The enrollment algorithm is based on the idea of evaluating a given reference set of feature vectors towards a transitivity-based quality criterion. We determine this criterion by computing a Transitivity Check Matrix (*TCM*) and checking it towards transitivity. If the criterion is not met, the algorithm returns that set member, that shows the greatest difference compared to all other remaining samples and replaces this reference with an additional handwriting sample, that has to be input by the user. This approach requires multi-dimensional reference feature vectors and can be parameterized regarding to number of reference set members and a termination border for the enrollment process. The algorithm is now described in detail.

3.1 Definitions

Let $S_i = (s_{i,1}, \ldots s_{i,NF})$ be the i-th sample feature vector member of the reference data set $R=(S_1, \ldots S_N)$, where NF is number of feature components that the sample consists of and N is the number of feature vector members that built a reference set.

Let $R'_i = R / S_i$ be the reference data set R excluding feature vector S_i. After a comparison between S_i and R, the distance vector $D(S_i,R)=(d_{SR1}, \ldots d_{SRNF})$ stores the similarity between a sample S_i and the reference set R, composed of distance values d_{SRj} for each $j \in [1..NF]$. Each d_{SRj} denotes the distance between feature vector component $s_{i,j}$. and the reference R as a numeric value. Then the Transitivity Check Matrix (*TCM*) is composed as follows:

$$TCM = \begin{pmatrix} D(S_1, R'_1) \\ \dots \\ D(S_N, R'_N) \end{pmatrix} = \begin{pmatrix} D_1 \\ \dots \\ D_N \end{pmatrix} \quad (1)$$

The sorted transitivity check matrix TCM_{Sort} is the matrix composed from the rows D'_1 .. D'_N of *TCM*, such that:

$d'_{1,SRj-1} \leq d'_{1,SRj}$ *for all* $j \in [2..N]$.

That is, the leftmost column in TCM_{Sort} shows values sorted in ascending order. Then, we define that the transitivity of enrollment *R* is approved, if and only if the following condition is fulfilled:

For all $k \in [1..N]$: $d'_{k,SRj-1} \leq d'_{k,SRj}$ *for all* $j \in [2..NF]$

That is, the row vectors on TCM_{Sort} are in transitive order.

3.2 Description of the Algorithm

The algorithm starts with an empty reference data set *R* and two parameters $N_{Samples}$ and N_{Max}. These parameters affect the number of samples in the reference set and the maximum number of data acquisition iterations until the algorithm aborts. The system then requests input samples from the user and collects them in the data set *R* until the set reaches the cardinality $N_{Samples}$. Now the sorted Transitivity Check Matrix TCM_{Sort} is computed from this reference and the transitivity check is performed. If this check returns a positive result, the enrollment *R* is accepted. Otherwise, the system checks the total number of already performed input iterations. If it exceeds N_{Max}, the algorithm terminates with an unsuccessful enrollment. Otherwise, the element sample reference S_i, which refers to the bottom row in TCM_{Sort}, is replaced by a new sample obtained from the user and the algorithm continues with computation of TCM*Sort*.

4. Test Results for the Enrollment Algorithm

Our tests have been performed on 3 different digitizer tablets (WACOM Intous A5, Acecat-III and Plawa Freestyler), 3 different test subjects and 5 different handwriting input semantics (Signature, user-defined pass phrase, user-defined symbol, fixed word, fixed numeric code). The reference feature vectors S_i were chosen two-dimensional, consisting of two discrete reference signals $x_i(t)$ and $y_i(t)$.

The reference was interpreted as the area in the plane, which is delimited by the minimum and maximum of each member signal value at each discrete time value. The distance function $D(S_i,R)$ was then defined as the cumulated quadratic distance the actual sample function values and the boundaries of the reference area.

The number of feature vector members in a reference $N_{Samples}$ was set to 4. Two enrollment strategies were tested, where Test A (Tab. 1) was based on an approach that requires re-acquisition of all 4 members of the reference set, if the reference set did not meet the TCM quality criteria. In Test B (Tab. 2) we applied the transitivity algorithm approach as described in this paper. Both tables present the number of enrollments and total number of samples that were requested for each user. All input samples were only recorded once and stored in a database during test A, and replayed in the same sequence during test B, providing identical test data for both tests. N_{FE} denotes the number of false enrollments for each subject, $N_{Samples}$ shows the total number of writing samples taken from each user and (*) mark cases, where the enrollment process was unsuccessfully terminated.

An important observation of these tests is the average number of samples that were required for a successful enrollment. While it was 7.84 in test A, it could be reduced to 6.62 by applying the transitivity algorithm in test B. Further, the 4 occurrences in test A, where the enrollment was aborted without success after 13-15 trials, could be reduced to 2 cases in test B.

Due to the limited number of only three test subjects, we cannot assume statistically safe conclusions. We are currently performing additional tests with up to 10 subjects and 3 different enrollment strategies. Actual test results can be referred to at [7].

5. Conclusion and Future Work

The first tests have proven that the general concept of our new transitivity strategy can significantly reduce FER errors of a signature verification system. The improvement can be noticed not only for signatures, but also for all types of handwriting semantics. However, the limited number of subjects does not allow safe statistical conclusions. In current tests, we are increasing the number of test persons to achieve statistical significance. Further we will evaluate the algorithm results for feature vectors with dimensions higher than 2 and for additional reference signals such as the pen velocity.

A very interesting aspect is our observation, that even FAR and FRR could be reduced by reference generation based on our presented scheme. It appears that the explicit replacement of weak feature vectors from the reference by additional samples leads to an overall quality improvement during the verification process. In future work, we will extend our evaluation by verification tests against reference sets built on different enrollment strategies. The actual test status can be found under [7].

References

[1] R. Plamondon, G. Lorette: "Automatic Signature Verification and Writer Identification – the State of the Art.", Pergamon Press plc., Pattern Recognition, 22, 2:107-131, August 1987

[2] R. Plamondon, F. Leclerc: "Automatic Signature Verification: The State of the Art – 1989-1993", International Journal of Pattern Recognition and Artificial Intelligence, 8:643-660, 1994

[3] Seong Hoon Kim, Myoung Soo Park, Jaihie Kim: "Applying Personal Weights to a Feature Set for On-line Signature Verification", IEEE International Conference on Document Analysis and Recognition, 882-885, Montreal, Canada, 1995

[4] C. Schmidt: „On-line Unterschriftenanalyse zur Benutzerverifikation", Shaker Verlag, ISBN 3-8265-6607-6, Aachen, Germany, 1999

[5] S. Hangai, S. Yamanaka, T. Hamanoto: "On-line signature verification based on altitude and direction of pen movement", IEEE Proceedings of the International Conference on Multimedia (ICME) 2000, 1:489-492, July 2000

[6] B. Wirtz: "Stroke-Based Time Warping for Signature Verification", IEEE Proceedings of the International Conference on Document Analysis and Recognition (ICDAR), pp. 179-182, 1995

[7] C. Vielhauer: "Advanced Test Results for Enrollment Strategies", www.platanista.com/icdar2001/testresults/

Semantic	ID	NFE Acecad	NSamples Acecad	NFE Freestyle	NSamples Freestyle	NFE Wacom	NSamples Wacom	NFE Average	NSamples Average
Signature	1	1	8	2	12	1	8	20	9,33
Signature	2	3	15(*)	1	8	1	8	13	8,00
Signature	3	1	8	1	8	1	8	19	8,00
Passphrase	1	1	8	1	8	1	8	19	8,00
Passphrase	2	1	8	1	8	1	8	19	8,00
Passphrase	3	1	8	0	4	2	12	15	8,00
Symbol	1	0	4	3	13 (*)	1	8	8	6,00
Symbol	2	0	4	1	8	1	8	14	6,67
Symbol	3	1	8	0	4	1	8	14	6,67
Fixed Word	1	1	8	1	8	0	4	18	6,67
Fixed Word	2	1	8	2	12	3	15 (*)	26	10,00
Fixed Word	3	1	8	1	8	2	12	20	9,33
Fixed Numeric Code	1	1	8	0	5	0	6	14	6,33
Fixed Numeric Code	2	0	4	0	8	0	14	12	8,67
Fixed Numeric Code	3	3	14 (*)	1	8	1	8	13	8,00

Tab. 1 Results of test A (4-again strategy)

Semantic	ID	NFE Acecad	NSamples Acecad	NFE Freestyle	NSamples Freestyle	NFE Wacom	NSamples Wacom	NFE Average	NSamples Average
Signature	1	0	5	0	10	0	5	20	6,67
Signature	2	1	15 (*)	0	6	0	8	7	7,00
Signature	3	0	6	0	5	0	8	11	6,33
Passphrase	1	0	5	0	5	0	7	10	5,67
Passphrase	2	0	5	0	7	0	6	12	6,00
Passphrase	3	0	5	0	4	0	11	9	6,67
Symbol	1	0	4	0	13	0	6	17	7,67
Symbol	2	0	4	0	5	0	6	9	5,00
Symbol	3	0	7	0	4	0	8	11	6,33
Fixed Word	1	0	5	0	5	0	4	10	4,67
Fixed Word	2	0	5	0	9	1	15 (*)	15	7,00
Fixed Word	3	0	5	0	5	0	9	10	6,33
Fixed Numeric Code	1	0	8	0	5	0	6	13	6,33
Fixed Numeric Code	2	0	4	0	8	0	14	12	8,67
Fixed Numeric Code	3	0	14	0	6	0	7	20	9,00

Tab. 2 Results of test B (Transitivity Strategy)