

Retail Applications of Signature Verification

Thomas G. Zimmerman¹, Gregory F. Russell², Andre Heilper³, Barton A. Smith¹, Jianying Hu²,
Dmitry Markman³, Jon E. Graham¹, Clemens Drews¹

¹IBM Almaden Research Center, 650 Harry Road San Jose California 95120;

²IBM T.J. Watson Research Center, Route 134 Yorktown Heights, NY, 10598;

³ IBM Haifa Research Lab, University Campus, Haifa, Israel 31905

Abstract

The dramatic rise in identity theft, the ever pressing need to provide convenience in checkout services to attract and retain loyal customers, and the growing use of multi-function signature captures devices in the retail sector provides favorable conditions for the deployment of dynamic signature verification (DSV) in retail settings. We report on the development of a DSV system to meet the needs of the retail sector. We currently have a database of approximately 10,000 signatures collected from 600 subjects and forgers. Previous work at IBM on DSV has been merged and extended to achieve robust performance on pen position data available from commercial point of sale hardware, achieving equal error rates on skilled forgeries and authentic signatures of 1.5% to 4%.

Keywords: Biometrics, dynamic signature verification, retail industry, identity theft

1. Introduction

There is a tradeoff between security and convenience. An unlocked car with the keys in the ignition is very convenient to use, but also easy to steal. Banks who issue credit cards want their customers to use their cards frequently, while minimizing their exposure to fraud. In general, once a credit card is issued, credit transaction practices lean toward giving the buyer the benefit of the doubt, as a moderate amount of fraud is preferable to lost business and frustrated customers that results if the authentication process is too burdensome or stringent. So, credit card use generally does not require photo ID; card possession and a modest attempt to approximate the signature on the card are typically sufficient to purchase merchandise.

To control fraud, banks run sophisticated software that analyzes card usage patterns to try and identify suspicious purchases. When suspicious card account activities are detected, the legitimate card owner is contacted by phone or card privileges are denied, which clearly has a negative impact on the legitimate card user. Determining when purchases are legitimate is an art, particularly during the holiday season when a significant portion of a retailer's income is generated in a few weeks of furious shopping.

In this paper we explore the use of dynamic signature verification as an appropriate biometric to reduce fraud while providing convenience for retail transactions. Biometric systems are concerned with identifying an individual based upon his or her distinguishing characteristics, which could be physiological or behavioral. Unlike static signature verification, dynamic signature verification uses not only the shape of an individual's signature, but actually analyzes pen motion timing captured during the signing process. Since signatures are currently the standard mechanism for a customer to authorize a credit card or check transaction, the growing deployment of signature capture stations provides much of the infrastructure necessary for applying dynamic signature verification to retail sales authentication.

2. Signature in Retail

Reliable authentication and authorization are increasingly becoming necessary for many commonplace activities such as boarding an aircraft, crossing international borders, entering a secure physical location, and performing financial transactions. Biometrics is a useful method to verify identity. Although face and fingerprint recognition have experienced wide attention due to the efforts of the Department of Homeland Security, these technologies have significant drawbacks in a retail environment.

Credit card transactions have used signatures for "authentication" at retail establishment for years. However, credit card security is surprisingly weak. Each time a customer makes a purchase at a retail store, the cashier may or may not

visually compare the offered signature against the reference on the back of the card. The primary purpose of the signature is to reinforce the cardholder's obligation to pay the bill. The retailer is obligated to keep the signed statement, or an electronic facsimile, to present to the customer if the charge is challenged. From a security vulnerability perspective, a signed card provides a potential forger a template to follow, while an unsigned card is an even greater invitation to fraud.

Many of the larger retail chains are choosing to install electronic signature capture devices to capture and store signatures. These multifunction devices can read magnetic cards, capture signatures and PIN numbers, and display advertisements [1]. Although their principle purpose is reducing paper handling and automating the transaction process, these devices are often suitable for collecting signatures for dynamic signature verification.

3. Identity Theft

Identity theft occurs when a thief assumes the identity of an individual, usually by collecting personal information on the victim, such as name, address, date of birth, Social Security number, and credit card number, and uses this information to bill charges to the victim's name. The Federal Trade Commission, which maintains records on identity theft activity in the U.S., estimates there were 10 million victims of identity theft in 2002, resulting in \$48 billion of losses to businesses and financial institutions, and \$5 billion in out-of-pocket expenses for consumer victims [2]. Half of the victims discovered the fraud by monitoring their accounts, another one-quarter were alerted by credit card issuers or banks who detected suspicious account activity.

The signature on the back of a credit card will not deter an identity thief who requests and signs a new card. However, by electronically monitor the accounts with dynamic signature verification algorithms, abrupt changes in signatures may be detected, indicating potential fraud.

4. Signature as a Biometric

A biometric can be classified as physiological or behavioral. Physiological biometrics measures some physical feature of the subject such as face, fingerprint, iris, hand and finger geometry. Behavioral biometrics measures a user actions, such as speaking, writing and walking. Most physical features remain relatively stable over time, while behavioral characteristics are in control of the subject and tend to change over the short and long terms due to health, physiological state and aging. Further, the subject can often create false negatives, hiding their true identity by consciously changing the behavior being measured. This implies that behavioral biometrics must be collected from a cooperative or unaware subject. While physiological biometrics may be adequately represented by a single sample, behavioral biometric generally requires several samples due to their inherent variability.

Signatures vary depending on fatigue, mental and physical state, and writing position (ergonomics). Herbst and Liu [3] discovered that pen accelerations, which are proportional to the muscle forces exerted by the signer, are consistent in a habitual signature. Natural pauses between sections of the name (interruption, reflecting on the sensation of writing on an unfamiliar surface), skipped strokes (eliminating a middle initial, replacing a first name with an initial), and decorative rubrics (dotting the "i" with a star, underlining the name) can effect the repeatability of signature samples. The signature signal consists of regions of high correlation of unknown duration separated by variable regions of low correlation. Therefore a verification strategy is to delineate regions (e.g. segmentation determined by pen down and pen up events), shift them individually to find the maximal of correlation to the reference signature(s), penalize shifting, and combine the results for an overall verification decision.

4.1. Static vs. Dynamic Signature Verification

When a subject signs on paper, they leave a static image of their signature. Forgers practice the art of reproducing the image (or shape) of a signature, with little regard to the motions that caused the image. Pen motions are ephemeral and are not captured on the paper. When a signature is captured with a digitizer, the pen motions (dynamics) are recorded. Commercial signature capture devices operate by recording pen position at a constant sampling rate, creating a vector file (typically 1 to 3 k bytes, uncompressed) that is more compact than storing the signature as an image. This is fortuitous for signature verification; from this set of position points, velocity, acceleration and other dynamic features can be derived.

When signing, the hand can operate in a regime known as ballistic motion, where the muscles are not controlled by sensory feedback. Ballistic motions are generally rapid, practiced motions whose accuracy increases with speed. Walking, playing piano, and golf swings are examples of ballistic motions, where the actions are based on prior repeated experience (training), rather than closed loop feedback. The individual muscle forces applied in signing center around 4 Hz, with a few subjects median signing frequency extending past 8 Hz, as show in Figure 1. The majority of the 250 subjects studied varies +/- 15% from their median frequency and half of their signatures sign within +/- 3% of their median frequency. When signing, the hand often moves faster than an individual could volitionally control through hand-muscle coordination. However this ballistic motion is repeatable, making a dynamic signature harder to forge than a static image and therefore a useful behavioral biometric.

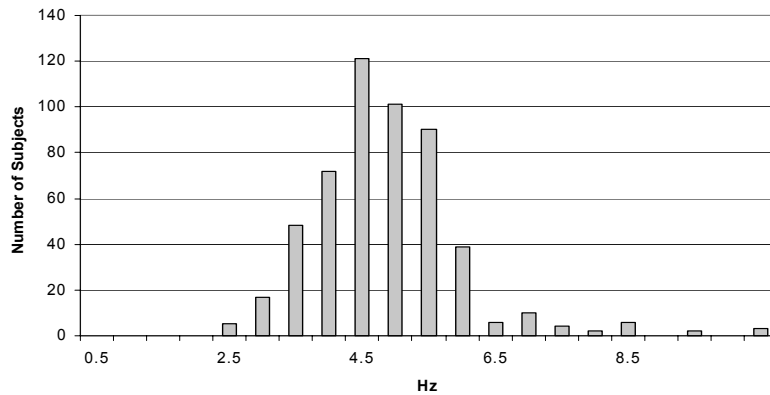


Fig. 1. Median Center Frequency when Signing

4.2. Identification vs. Verification

Identification answers the question “who are you”, while verification answers the questions “are you who you claim to be”. Identification is a 1:N match, which can be quite difficult when N is in the millions, for example potential shoppers at a large retailer. Verification is a much more tractable problem, requiring a 1:1 match, based on a reference biometric retrieved by an identification claim (typically from the user). The claim need not be secure or secret. The user can enter their phone number or other information from memory, or present a token as an identification claim, for example a loyalty card with a customer number encoded in a bar code or magnetic stripe. The security is not in the identification claim; rather it is in the reference biometric pointed to by the claim and the confidence of the match between the reference biometric and the offered biometric. The identification claim turns an identification problem into a verification problem, greatly simplifying the search and matching task.

Since verification checks an offered biometric to a reference biometric, it scales well for commercial systems with a large enrollment population. While identification is required for security applications like watch lists where subjects are tested against a list of “people of interest”, in this paper we are interested in verifying the identity of individuals for purposes of conducting financial transactions in a retail store environment (e.g. grocery, clothing, or electronic merchandise stores).

5. DSV System Elements

5.1. Dynamic Signature Verification Objective

The challenge to a Dynamic Signature Verification system can be states as follows; given a small set of reference signatures, determine the probability that a new signature submitted for testing is authentic. From this probability a binary (Yes/No) verdict results typically by applying an acceptance threshold to a match probability. If the acceptance threshold is too high, authentic signers may be rejected (False Reject), annoying customers. If the acceptance threshold is set too low, forgers will have an easier time passing as an authentic signer (False Accept). As a performance biometric, every instance of a signature is different, and a writer’s consistency is influenced by many factors. Some people are good mimics, and given enough practice and information, can produce very good imitations of authentic

signatures. Setting the acceptance is a delicate balance of not annoying too many customers with False Rejects, while maintaining enough security to reject all but the most skilled, practiced, or lucky forgers (e.g. picking a signature that is easy to forge).

5.2. Signature Acquisition

Signature capture devices operate with a common principle of capturing the X, Y position of the pen at a constant sample rate. The least expensive devices use a resistive membrane that, when pressed together by the pen, create two resistive paths, proportional to location [4]. More accurate, albeit expensive systems use inductive coupling from a coil in the pen to an array of coils in the pad to capture pen position [5], and often pen tilt and pressure as well. In retail applications the main purpose of the capture station is to save and retrieve a reasonable facsimile of the image of the signature. Hence pressure and tilt are not needed, and jitter in the sampling rate and non-linearity in the position measurement are tolerated. To save on storage and transmission requirements, redundant position data points are often thrown away. While this has little effect on the resulting image, it causes great disturbances to signature verification algorithms that rely on the constant sample rate to extract velocity and acceleration information. Fortunately on the more sophisticated capture devices used for retail [1], actions like throwing away points is controlled by firmware that can be altered in the field. In our work we have limited our verification to position data, assuming that tilt and pressure would not be available on all retail systems.

The signature capture devices are mounted in a plastic case often with a slight (e.g. 15 degrees) tilt to make signing more comfortable. Signature is based on the movement of muscles and is affected by the ergonomics of the mounting of the capture device. Placing a capture device vertically, as observed on some self-check out stations, would no doubt require the customer to sign with different muscles than when signing on a flat surface, potentially degrading the performance of a verification system. Slippery writing surfaces and poor or delayed feedback can also upset the signature process and reduce reliability. Ideally customers are enrolled and verified on the same hardware, with the same physical layout and arrangement. Signature files may be re-sampled to compensate for capture devices with different spatial and temporal sample rates, but such simple transformations are not possible for different ergonomics.

5.3. Feature extraction

It is possible to compare signatures by extracting a set of approximately invariant metrics of two signatures, then comparing these two sets, e.g. by computing the Euclidean distance between them. Such metrics may include, for example, signing time, aspect ratio, moments of inertia, or frequency spectra. As these are fairly blunt metrics, it is possible to find signatures that look very dissimilar, yet provide fairly good matches.

Worthington and Chainer [6] used spectral features together with correlation measurements to compute distance metrics between signatures. With the richness of the acceleration data from the signature capture pen, this approach was very successful. The correlation measurements were computed on patches of signature of about 700 millisecond duration, allowing for some variability in signing speed, while penalizing speed variations within the patches. We have retained the correlation metric of [6], but combined it with additional features.

Features may also be extracted from the signatures differentially. We use dynamic time warping [7] to find corresponding points in two signatures, and then compute features based on the detailed differences between corresponding regions. This is potentially much more powerful, particularly if the information about the detailed alignment and its cost is retained and used in the classification step, as our DSV engine does.

5.4. Signature alignment with reference

Dynamic Time Warping (DTW) compensates for minor signature speed variations, normalizing the time duration of signatures, to provide correspondence between points in offered signatures and points in reference signature. The algorithm identifies the best monotonic map between points in the test signature and points in the reference signature, based on some cost function. The result may be interpreted as a monotonic curve in a coordinate system in which one axis is the index of the points in the reference, and the other axis is the index of points in the test signature. The process also results in a cost value, which reflects the quality of the alignment and the similarity between corresponding portions of the signature.

Figure 2 shows the alignment obtained from our DTW engine for two artificial signatures from the same writer in which the first names are different, but similar and the last names are the same. The robustness of the algorithm is evident in the plausible alignment of various portions of the signature, in spite of the fact that many parts do not correspond well.

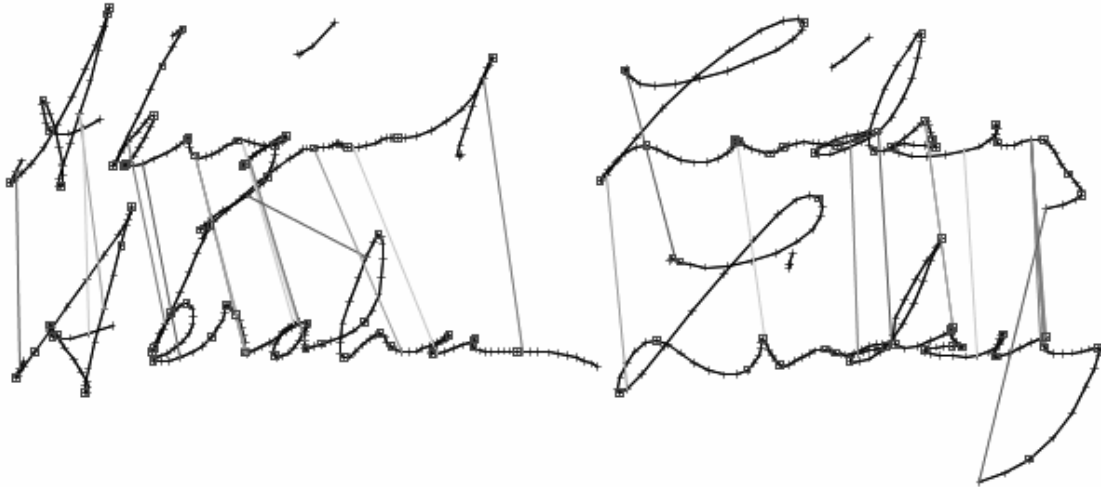


Fig. 2. Matching Segments in a Signature Sample

5.5. Classification

The system uses a classifier based on separate Mixture of Gaussians models for authentic and forged signatures. The feature extraction is effective enough to achieve good performance modeling only three dimensions. We have also evaluated Fisher Discriminant classifiers (in higher dimensions), Support Vector Machines, and neural networks. The latter two achieve performance similar to the Mixture of Gaussians model, but the simplicity of the Mixture of Gaussians provides us with much greater flexibility in tuning the system.

The system achieves good modeling precision with 3 to 4 Gaussians in each mixture. We get good results with either full covariance modeling or by applying Principal Component Analysis (PCA) to the data before modeling and constraining the models to diagonal covariance matrices.

6. Performance Metrics

The verification process accepts or rejects the identity claim of the subject based on the match between the offered and reference biometric. There are two types of errors in verification: false accepts (FA), which indicates the likelihood that someone may be falsely accepted by the system, and false rejects (FR), which indicates the likelihood that a genuine user may be rejected by the system. These measures are expressed in percentage (of error) terms; the lower the percentage, the better the system performance. Most verification systems have an acceptance threshold allowing the trade-off of these errors. By lowering the acceptance tolerance fewer valid customers are rejected (lower FR), but more imposters are accepted (higher FA). The threshold allows system users to optimize the errors for their needs, for example setting higher acceptance thresholds for higher value transactions.

Setting the threshold so the FA equals FR achieves a condition known as equal error rate (EER), a convenient method to compare verification algorithms. But comparisons based solely on EER (assuming all other comparison conditions are equal), obscures the impact of varying the threshold. For example two systems with the same EER, may exhibit different false rejects and false accepts when the threshold is changed.

Most signature verification applications allow multiple attempts to verify, providing greater robustness against spurious problems. The system can be characterized either by its raw FA and FR rates, or by the composite or “session” FA and FR rates. Better error rates may be obtained for sessions than for single signatures, by appropriately adjusting the thresholds for first, second, or additional attempts. We report single signature FA and FR rates. Estimates may be

extrapolated for session error rates, but the actual performance is influenced by the specific protocol and system dialog with the user.

Other factors often ignored in the marketing of biometrics are the failure to enroll (FTE) and failure to acquire (FTA). Not all people exhibit the biometric that can be sufficiently and repeatedly captured. FTEs occur with subjects who cannot produce biometric samples with enough detail or repeatability to be discriminated or matched. The enrollment must be rejected, as the subject's biometric does not allow reasonable false reject or false accept rates.

FTAs are verification attempts that fail because the biometric acquisition does not yield a sample of sufficient quality to be used by the verification system. Failure to acquire may be attributed to system malfunctions, or to physical anomalies at time of capture, e.g. soft or greasy finger for fingerprint readers. In the case of fingerprint or iris scanning, the problem is often evident in the data preprocessing, when it may be impossible to extract adequate minutia, or an iris may not be found anywhere in the image. In the signature domain, any but the most obvious failures may be hard to identify. An FTA would result if the collection is terminated without any ink being drawn, or if a single dot is collected. If the user writes a big squiggle that doesn't resemble the signature at all, it may just be identified as a rejection, rather than as an FTA.

6.1. Influence of data collection

The art of fingerprint and face recognition has dramatically improved in the past few years due to heightened interest in national security and standardized methods of testing, primarily resulting from large biometric databases and testing performed by government bodies [8]. Interest in Dynamic Signature Verification is increasing due to the explosive growth of identity theft and the desire in retail to increase the use of electronic tender (due to lower transaction costs attributed to automated clearing mechanisms). However the lack of public databases makes comparison of DSV algorithms difficult. While each vendor may test their DSV on their private database, these results do not allow comparisons across vendors, since data sets and the testing conditions may vary dramatically. Factors such as capture hardware, failure to enroll, variations in signature acquisition configurations, and training procedures, all impact the results. Systems that use more degrees of freedom (position, pressure, and tilt) tend to perform better than system that captures fewer signature attributes.

6.2. Signature Corpus

Our training and testing databases consist of three sets of data collected in 1994, partly in the U.S. and partly in Israel, and four sets of data collected more recently, on better collection devices, again in both the U.S. and Israel. We have a mix of enrollments, some having more or fewer verification attempts, and some having more or fewer forgery attempts. The degree of practice of the forgers also varies. Unfortunately the protocol used in collecting the 1994 data is unknown. There is evidence of some mislabeling of the 1994 data as well, so this data is used exclusively for training, and the more recent data is used for training or evaluating.

We have observed strong evidence that signature styles differ between the U.S. and Israel in a manner that significantly affects verification accuracy. Systems trained on U.S. signatures perform poorly when tested on Israeli signatures, and vice-versa. The composition of Israeli signatures is primarily a multitude of straight lines, while US signatures tend to be a few long strokes with lots of curves. We have concentrated on U.S. trained systems, and do not report on the Israeli data at this time.

7. Signature Parameters

Signature is a personal behavior developed by the individual that is surprisingly consistent. Measurable features that are consistent for a signer are useful for filtering verification candidates before more calculation intensive operations are performed. We collected at least ten signatures from 250 subjects using a Wacom Intuos II digitizer [5]. From that data we measured several features. The average signing time centers around 3.2 seconds, as shown in Figure 3, with a few extreme subjects signing in one second and eight seconds, respectively. The signing times for the subjects varies about +/- 30%. Taking a closer look, we see in Figure 4 that some subjects have a wide range in signing times. The extreme examples are possibly due to interruptions during signing, such as a purse falling off the shoulder of a signer, or someone talking to the subject.

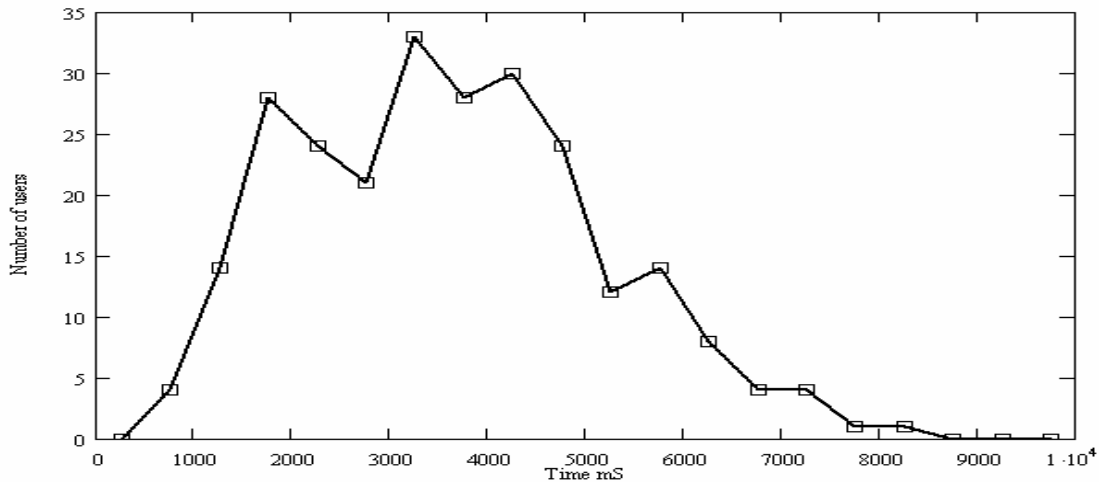


Fig. 3. Signing Times for Each Subject

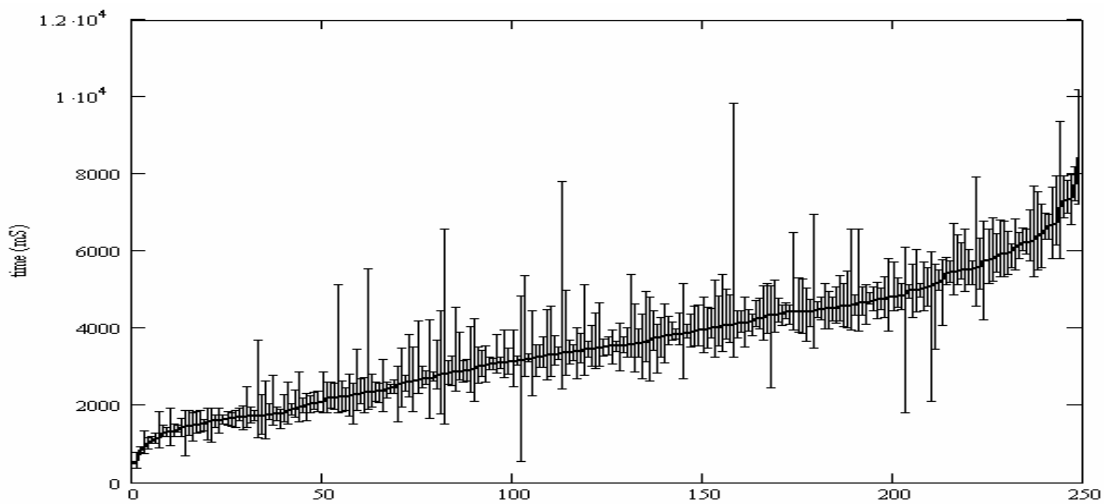


Fig. 4. Signing Times for Each Subject

7.1. Forgeries

A biometric database of authentic signatures is useful in demonstrating false reject rates. Authentic signatures may also be used to compute random forgery detection rates, but this is a rather weak assessment of system security.

Random forgery is the ability of the system to distinguish between authentic signatures from different people. No attempt is made to forge signatures – each authentic signature is merely tested against the enrollments of all other subjects. Ideally each signer’s signature should be unique and not match the signature of any the other subject in the database.

Running a random forgery test is a good way to detect any severe system weaknesses. Systems trained exclusively on good forgeries may be susceptible to oddly simple attacks, which may be revealed by running a random forgery test. DSV systems should be expected to have very low random forgery false accept rates, typically no more than a small fraction of a percent. Accurate numbers require very large numbers of subjects, as large numbers of signatures from a small set of subjects is unlikely to expose anything other than system bugs.

To evaluate true false accept rates, skilled forgeries are required. For biometrics in general, and for signatures in particular, skilled forgery performance is difficult to compare across systems, due to the combination of serious privacy concerns and the strong variability of the skill of forgers and protocol used to collect forgeries. Do the forgers see the authentic specimen? Do they get a feedback about how well they do? How motivated are they? How long do they practice? We have collected such a forgery database for development use, and can report some performance metrics, though they are of limited usefulness. There is no publicly available set of signatures available yet for comparative studies between DSV systems.

8 Performance

Figure 5 shows the raw, single signature FA / FR curve for our DSV on the forgery database described above, with an equal error rate of 1.7%. This version of the system uses a simple 6 signature enrollment. Rolling enrollment functionality is currently being implemented in the system, and performance numbers are not yet available.

A rough estimate of two-signature session rates can be made from the single signature data. With the threshold set to produce single signature FA rate of 0.24 %, the corresponding FR rate is approximately 4%. Based on the observed scores for authentic signatures, this would result in session FA rate of approximately 0.4%, and session FR rate of approximately 2.2%. This is dominated by a single individual with a raw false reject rate of 66%, probably due to an enrollment problem. In a retail setting, this might require a re-enrollment, or one or two ID checks, after which the user would probably return to 5% raw FR rate, and the aggregate performance would be approximately 0.55% session FR rate. More accurate estimates of actual session performance would require a pilot in a more realistic environment.

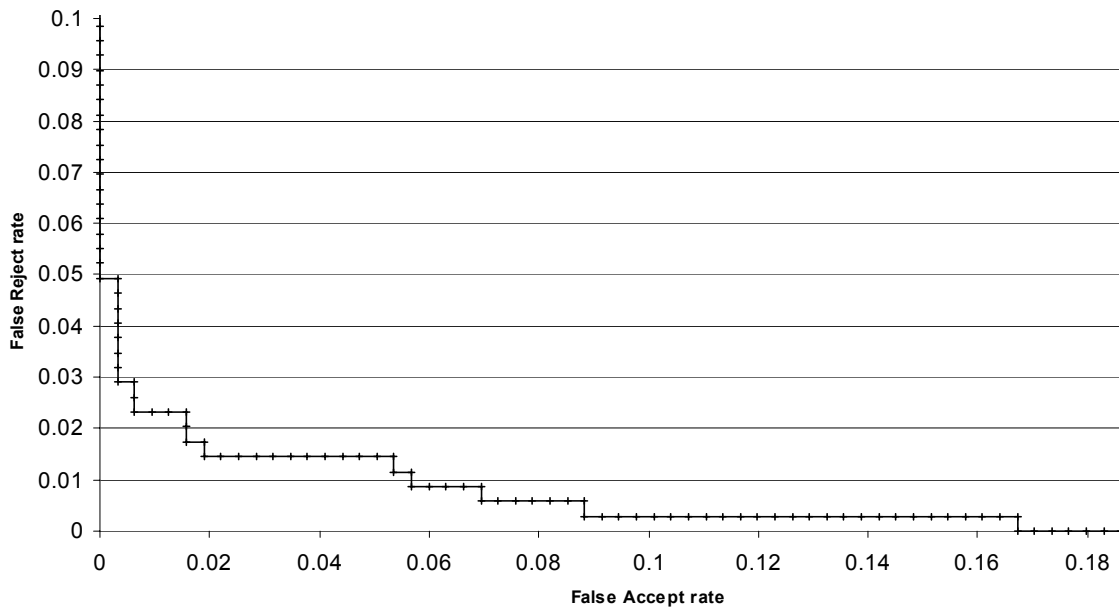


Fig.5. False Accepts vs. False Rejects Rate for 27 Forgers

9 Conclusions

We have developed a dynamic signature verification system to meet the needs of the retail sector. Our DSV engine merges and extends previous work at IBM. Tests conducted on skilled forgeries and authentic signatures produced equal error rates of 1.5% to 4%. The behavioral nature of signature biometrics requires careful attention to capture device ergonomics, enrollment and verification procedures. To be valid, comparison of DSV system should use the same enrollment and forgery corpus and conditions, and will provide realistic results when large databases of signatures become available.

Acknowledgements

The authors would like to thank Tim Chainer for his expert advice on dynamic verification techniques, Jane Snowdon for her editorial assistance, the IBM First Of A Kind (FOAK) program for funding this research, and the hundreds of volunteers at IBM Almaden, T.J. Watson, and Haifa Research Laboratories for donating their biometrics to the IBM signature corpus.

REFERENCES

1. http://www.ingenico.com/download/pdf/entouch1000_uk.pdf *eN-Touch 1000 Touch Screen Terminal For Signature Capture Technical Specification* Ingenico Corporation
2. <http://www.ftc.gov/opa/2003/09/idtheft.htm> *FTC Releases Survey of Identity Theft in U.S. 27.3 Million Victims in Past 5 Years, Billions in Losses for Businesses and Consumers* Federal Trade Commission, Press Release, September 3, 2003
3. Herbst, N.M. and Liu, C.N., *Automatic Signature Verification Based on Accelerometry*, IBM Journal of Research and Development Vol 21, No. 3 pp. 245-253, May 1977
4. <http://www.epadlink.com/> *ePad POS Electronic Signature Solution* Interlink Electronics
5. <http://www.wacom.com/productinfo/4x5.cfm> *Wacom Intuos2 4x5 Product Information*, Wacom Technology Corporation
6. Chainer, T.J., and Worthington, T.K. *Segmentation Algorithm for Signature Verification*, United States Patent 4,553,258, November 12, 1985, assigned to International Business Machines, Armonk, New York.
7. Kruskal, J.B. and Liberman, M. (1999) *The symmetric time-warping problem: from continuous to discrete*. In Sankoff, D. and Kruskal, J. (eds), *Time Warps, String Edits, and Macromolecules: The Theory and Practice of Sequence Comparison*. CSLI Publications, Stanford, pp. 125-161.
8. <http://www.nist.gov/srd/biomet.htm> *Scientific and Technical Databases-- Biometrics*, National Institute of Standards and Technology