Global feature selection for on-line signature verification

Hamed Ketabdar [1] , Jonas Richiardi and Andrzej Drygajlo

Swiss Federal Institute of Technology Lausanne (EPFL)
Signal Processing Institute
Perceptual Artifical Intelligence Laboratory

Abstract. A large number of features can be used to represent on-line handwritten signatures in verification tasks. Depending on the signature database and acquisition conditions, some features will not help in separating writers in the feature space so that an appropriate decision boundary will be hard to estimate. Other features will provide good separability between legitimate system users and their forgers. This paper proposes a signature feature selection algorithm combining a modified Fisher ratio cost function and a sub-optimal but fast search method to explore an initial feature space of candidate global features. A large number of candidate feature subsets of various sizes is evaluated, and it is shown that our modified Fisher ratio correlates highly with experimental verification error rates. The need for forgery data in the feature selection phase of signature verification systems development is also investigated, and we postulate that user-to-user separation is a good indication of user-to-forger separation.

Keywords: feature selection, separability measure, discriminant analysis.

1. Introduction

Over the years, many features have been proposed to represent signatures in verification tasks. We distinguish between *local* features, where one feature is extracted for each sample point in the input domain, *global* features, where one feature is extracted for a whole signature, based on all sample points in the input domain, and *segmental* features, where the signature is subdivided into segments (typically based on velocity) and one feature is extracted for each segment. This paper focuses on global features.

Signature verification can be considered as a two-class pattern recognition problem, where the authentic user is a class and all her forgers are the second class. Feature selection refers to the process by which descriptors (features) extracted from the input-domain data are selected to provide maximal discrimination capability between classes. In previous work on feature selection for signature verification, statistical methods such as Linear Discriminant Analysis (LDA) have been applied for segmental features to obtain the discriminative power of each individual feature [2]. In [5], a statistical distance measure (feature-by-feature difference of means between two users scaled by standard deviation) is used to select the best feature subset out of a 42-features and a 49-features candidate list. In [1], a backwards search procedure starting from 44 global features is used with an equal error rate (EER) cost function to select a subset of features. Selection of local features based on classifier score (match distance called dissimilarity measure) is performed in [4]. Recently, a mix of 22 local and global features extracted from the SVC 2004 database were extracted and ranked individually by a “consistency” measure, essentially a difference of distance measure-specific means scaled by the standard deviations [6].

For our research, we gathered a large number of global features (more than 150 extracted from 60 papers dating from 1983 to the present) and we use a near-optimal feature space search algorithm (the floating search) along with an improved version of the Fisher ratio as a cost function. In order to take into account effects of correlation between feature vector components, the cost is computed on whole feature vectors instead of individual features. Because class separability depends on the classifier used, the best feature vector will not be the same for different classifiers [3]. Therefore, we are narrowing the problem down to finding the optimal feature vector with respect to a Bayes classifier.

Our approach works as follows: a series of search steps (described in Section 3.) starting from an initial set of features is used. At each search step a cost function (described in Section 2.) measuring the discriminative ability of the feature subset is applied. We then correlate cost function value with the equal error rate (EER) of a Gaussian mixture model (GMM) classifier for signature verification (Section 4.).

2. Measuring discriminative ability of feature subsets

In order to evaluate a candidate feature vector subset at each search step, it is necessary to define an objective measure or cost function. The measure should be high when classes are more easily separable in feature space. Many types of cost functions can be used for feature selection, encompassing distance measures (e.g. Euclidean distance, Divergence, Bhattacharyya distance), information content measures (e.g. Mutual information [10]), and error-rate measures, where the error rate of the classifier is used directly as a criterion to evaluate the current subset (wrapper method).

---
[1] now with IDIAP Research Institute, Switzerland

For this work, we use a modified version of the Fisher ratio as a cost function, which we explain below.
The Fisher ratio provides a good mathematical framework for expressing the idea that within-class variability should be small, while between-class variability should be large. The within-class scatter matrix is defined as follows:

$$\mathbf{S}_w = \sum^{M} P_m \mathbf{\Sigma}_m, \tag{1}$$

where M is the number of classes, P_m is the prior probability of class m, and $\mathbf{\Sigma}_m$ is an estimate of the covariance matrix. The class covariance matrix is computed as:

$$\mathbf{\Sigma}_m = E[(\mathbf{x} - \mu_m)(\mathbf{x} - \mu_m)^T], \tag{2}$$

where μ_m is the mean of class m features. The between-class scatter matrix is defined as:

$$\mathbf{S}_b = \sum^{M} P_m(\mu_m - \mu_0)(\mu_m - \mu_0)^T, \tag{3}$$

where μ_0 is the global mean vector, computed over all classes as follows:

$$\mu_0 = \sum^{M} P_m \mu_m, \tag{4}$$

From Eqs. 1 and 3, one computation commonly used for the "classical" Fisher ratio is:

$$J = \frac{trace(\mathbf{S}_b)}{trace(\mathbf{S}_w)} \tag{5}$$

Thus, J will be large when class samples (signature presentations) are narrowly clustered around their class means and class clusters are well separated. However, one problem with this definition is that the between-class separation is only measured with respect to the global mean. Therefore, if the weighted sum of individual class distances to the global mean (trace of $\mathbf{S}_b$) stays the same, J will not become larger for classes that are pairwise further apart. To correct this problem, we add an Euclidean distance term E, representing the averaged two-by-two distance between class means, to the J criterion (of which we take the root because it represents a ratio of squared distances):

$$E = \frac{1}{M(M-1)/2} \sum_{i \neq j} dist(\mu_i, \mu_j) \tag{6}$$

$$J' = \sqrt{J} + E \tag{7}$$

2.1 What definition of class should we use?

Signature verification is a unique biometric modality because it is the only one to be systematically tested against dedicated impostors instead of random impostors. Thus, while in modalities such as face verification the class definitions are clear cut (user's face versus all other faces), in signature verification we have two possibilities.
The first possibility is to consider each user as a class, and to compute the modified Fisher ratio J' of equation 7 using these classes (as many classes as users). This will measure the separability between authentic users, which is equivalent to measuring the separation ability of a feature subset for random impostors. We denote this definition of our measure J_ω.
The second possibility is to define two classes for each user: one authentic class and one forgery class. In this case the modified Fisher ratio J' is computed, for each user, with these two classes. Then, the mean over all the users is computed to produce what we call the "authentic-forgery" cost function $J_{\bar{\omega}}$.

3. Search method

With a large number of initial features, exhaustive search of the feature subset space becomes computationally intractable, as an initial set of F features would result in $2^F - 1$ possible combinations. Therefore, we use floating search [8], a suboptimal search strategy which affords the flexibility of reconsidering a previously discarded feature or to remove a selected feature. Once the floating search has reduced the initial feature set to a more tractable dimension, optimal (exhaustive) search can be performed on the reduced space of potential feature subsets.

4. Experiments

The database used for the tests is a 25-users subset of the MCYT database [7], which provides (x, y) coordinates, pressure, azimuth and elevation data sampled from a Wacom Intuos2 tablet at 100Hz. Each user has 25 authentic signatures, and a total of 25 forgeries collected from 5 different forgers. The forgers were allowed to practice forgery as they needed and were provided with a static representation of the signature. Preprocessing is done by translating each signature to start at coordinates $(0, 0)$, but no rescaling or rotation takes place because the acquisition was done using an inking pen on a strict paper grid providing immediate feedback and orientation information. The 46 global features extracted for these experiments were selected from our list of more than 150 features for their representativity and frequent use in signature verification litterature. They are shown in Table 1.

number of samples (T)	signature height (H)	signature width (W)
H to W ratio	T to W ratio	avg. velocity
max velocity	avg. velocity $\div$ max velocity	avg. x velocity
var. of x velocity	num. pts. with positive x velocity	RMS velocity
var. of velocity	pen down samples (T_d)	time of max velocity $\div T_d$
time of max x velocity $\div T_d$	RMS acceleration	avg. acceleration
var. of acceleration	avg. pressure	max pressure
point of max pressure	avg. azimuth	avg. elevation
avg. y velocity	x y velocity correlation	first moment
max pressure-min pressure	max x velocity	avg. x acceleration
max y velocity	avg. y acceleration	var. of pressure
point max. velocity $\div T_d$	num. points with negative x or y velocity $\div T_d$	
max. acceleration	num. points with positive x or y velocity $\div T_d$	
tangent histogram in 8 quadrants: $S_q = card\left\{\theta_t : (q-1)\frac{\pi}{8} < \theta_t < q\frac{\pi}{8}\right\} \div (T-1)$ where $t = 2, \ldots, T$ and $q = 1, \ldots, 8$		

Table 1

Initial set of features

4.1 J_ω criterion-error rate correlation

To test whether the J_ω value is a good predictor of classification performance, these global features are used as an input to the floating search algorithm, which produces a secondary set of 12 features shown in Table 2. The J_ω value is computed for all of the 4095 resulting subsets. Then, these feature subsets are used with a GMM classifier [9] to obtain corresponding EER figures. The classifier uses 5 training signatures, and models users using 2 diagonal covariance matrix Gaussian components per user model. Score normalisation is performed for each user by a cohort background model of 6 Gaussian components with diagonal covariance matrix trained from the pooled training data of all other users.

As shown in Fig. 1(a), the J_ω criterion is highly correlated with classifier performance at EER, with a Pearson linear correlation coefficient [3] of -0.64. This means that significant compute time savings can be achieved by using this criterion instead of direct EER measure for feature selection. However, as all filter methods of feature selection, the compute time saving comes at the expense of possibly missing the best feature subset, which would be provided by wrapper methods.

The Fisher ratio in itself would be a poor feature selection criterion for on-line signature verification, as can be seen on Fig. 1(b). This is likely due to the fact that high-dimensional covariance matrice have a large number of free parameters and that global features only provide one point of data per signature realisation. Thus, it is likely that covariance estimates are severely biased. Furthermore, since many global features are not normally an unimodally distributed within their classes, a single covariance matrix may not be an adequate representation of their distribution.

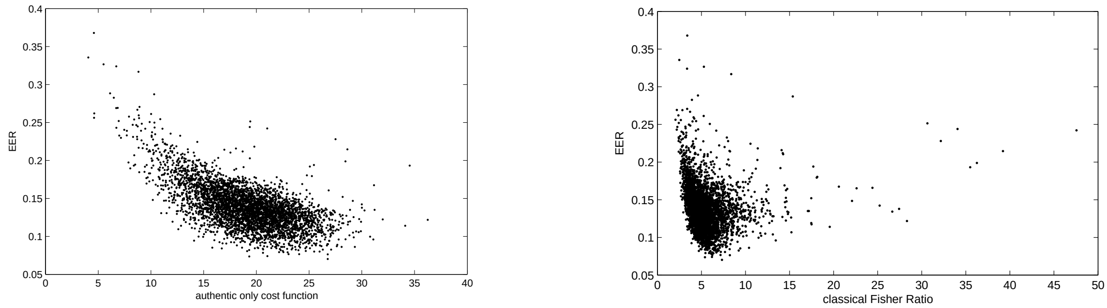
4.2 Need for forgery data

For real deployments of signature verification systems, forgery data may be difficult to collect. Therefore, we tested whether a feature set providing large user-to-user separation (high J_ω) would also provide a large authentic-to-forger separation (high $J_{\bar{\omega}}$). This was done by computing $J_{\bar{\omega}}$ in addition to J_ω using the same procedure as before. Then, the Pearson linear correlation coefficient was computed and found to be 0.78. This figure and the

1. T	2. avg. velocity	3. avg. velocity $\div$ max velocity
4. num. points with +ve x velocity	5. T_d	6. avg. pressure
7. variance of pressure	8. max. pressure	9. point of max. pressure
10. avg. elevation	11. S_1	12. num. points with +ve y velocity $\div T_d$

Table 2

floating search results: secondary set of features



(a) Correlation between EER and J_ω value ($\rho = -0.64$)

(b) Correlation between EER and classical Fisher ratio ($\rho = -0.1$)

Fig. 1. Assessment of predictive ability of the J_ω criterion

scatterplot shown in Fig. 2 suggest that user-to-user separation measure J_ω is a good predictor of authentic-to-forger separation measure $J_{\bar{\omega}}$ for GMM classifiers, and means that verification systems may be in a developed and optimised without the need for forgery data. It should be noted that forgery data in the MCYT database is only "skilled" in the sense that forgers were allowed to practice on a static image of their target, and that results may be different where over-the-shoulder forgeries are used.

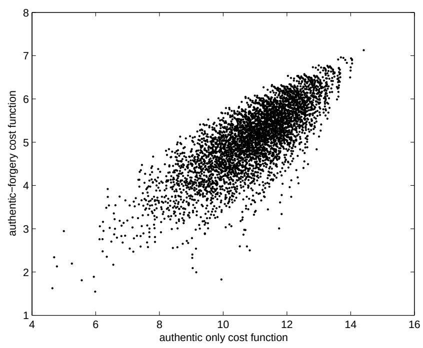


Fig. 2. Correlation between J_ω and $J_{\bar{\omega}}$ cost functions ($\rho = 0.78$)

4.3 Error rates

Lastly, the best feature set (highest J_ω criterion value) found by the floating search procedure and consisting of the 12 features in Table 2 is tested in a verification task. The classifier used is the same as in the previous experiment, and 50 users (each providing 5 training signatures) are used for testing, resulting in 1000 genuine signatures and 1250 forgeries being tested. As can be seen on the Detection-Error Trade-off (DET) curve of Fig. 3, the EER is around 4.5%, a reasonable figure considering that each user is modelled using a total of 5 12-dimensional training vectors. It is probably the case that a GMM is not the most appropriate classifier to use with such a limited amount of training data (the parameter estimates will likely be too biased), and nonparametric models could be used instead.

5. Conclusions

We have presented a methodology for selecting a subset of global features out of a large initial set of features. We proposed a modification to the classical Fisher ratio and obtained higher correlation with classifier performance for a signature verification task. Our criterion is effective for measuring the class separation ability provided by

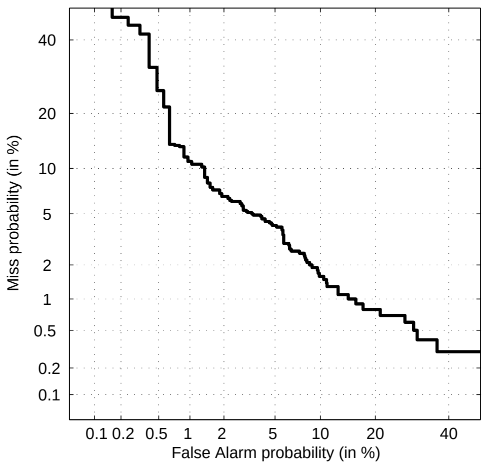


Fig. 3. DET curve for best global feature subset (4.4% EER on 50 users)

feature vectors. We also showed two possible ways of computing distance between classes, one in which only authentic signatures are used, and a second in which both authentic and forgery data is used. We then showed that the two measures where highly correlated, and postulated that the unavailability of forgery data may not be a major drawback in signature verification system design; separation of legitimate users in feature space is a good indication of the distance between forgers and their targets.

Further work will include exploring other separability measures such as joint mutual information, testing with non-parametric classifiers and using a larger signature database.

Acknowledgements

We thank J. Ortega-Garcia and J. Fierrez-Aguilar for providing the MCYT subcorpus, as well as the anonymous reviewers for their useful comments.

References

[1] H.D. Crane and J.S. Ostrem. Automatic signature verification using a three-axis force-sensitive pen. *IEEE Transactions on Systems, Man, and Cybernetics*, 13(3):329–337, June 1983.

[2] J.G.A. Dolfing, E.H.L Aarts, and J.J.G.M. van Oosterhout. On-line signature verification with Hidden Markov Models. In *Proc. International Conference on Pattern Recognition 1998*, volume 2, pages 1309–1312, Aug. 1998.

[3] K. Fukunaga. *Introduction to statistical pattern recognition*. Academic Press, second edition, 1990.

[4] A.K. Jain, F.D. Griess, and S.D. Connell. On-line signature verification. *Pattern Recognition*, 35:2963–2972, 2002.

[5] Luan L. Lee, Toby Berger, and Erez Aviczer. Reliable on-line human signature verification systems. *IEEE Transactions on pattern analysis and machine intelligence*, 18(6):643–647, June 1996.

[6] H. Lei and V. Govindaraju. A study on the consistency of features for on-line signature verification. In *Proc. Structural, Syntactic, and Statistical Pattern Recognition: Joint IAPR International workshops, SSPR 2004 and SPR 2004*, pages 444–451, Lisbon, Portugal, 2004.

[7] J. Ortega-Garcia, J. Fierrez-Aguilar, D. Simon, J. Gonzalez, M. Faundez-Zanuy, V. Espinosa, A. Satue, I. Hernaez, J.-J. Igarza, C. Vivaracho, C. Escudero, and Q.-I. Moro. MCYT baseline corpus: a bimodal biometric database. *IEE Proc. Vision, Image and Signal Processing*, 150(6):395–401, December 2003.

[8] P. Pudil, J. Novovicova, and J. Kittler. Floating search methods in feature selection. *Pattern Recognition Letters*, 15:1119–1125, November 1994.

[9] J. Richiardi and A. Drygajlo. Gaussian mixture models for on-line signature verification. In *Intl Multimedia Conf., Proc. 2003 ACM SIGMM workshop on Biometrics methods and applications*, pages 115–122, Berkeley, USA, Nov. 2003.

[10] G.D. Tourassi, E.D. Frederick, M.K. Markey, and C.E. Floyd. Application of the mutual information criterion for feature selection in computer-aided diagnosis. *Medical Physics*, 28(12):2394–2402, December 2001.