Learning Sparse Kernel Machines with Biometric Similarity Functions for Identity Recognition

Battista Biggio, Giorgio Fumera, and Fabio Roli
Department of Electrical and Electronic Engineering, University of Cagliari
Piazza d'Armi, 09123 Cagliari, Italy
{battista.biggio,fumera,roli}@diee.unica.it

Abstract

We investigate the application of similarity-based classification to biometric recognition, interpreting similarity functions used in biometric systems (i.e., matching algorithms) as kernel functions. This leads us to formulate biometric recognition as a distinct two-class classification problem for each client, which can be solved even when no representation of biometric samples in a feature space of fixed dimensionality is available. We discuss the relationship of our approach with cohort-based methods, and show that using support vector machines exhibits several advantages, in terms of the automatic selection of the cohort size and elements, and of the possible update of each user model. A biometric verification setting is considered for the formulation of the approach, but experimental results with face and fingerprint data sets are reported for both verification and identification settings.

1. Introduction

Biometric verification is usually formulated as a hypothesis testing problem. Let $T = \{t_1, \ldots, t_N\}$ denote the set of templates (user models) of N enrolled clients (e.g., N template face images); $\Omega = \{\omega_i\}_{i=1}^N$, the set of the corresponding identities; x, the sample submitted by a user at verification phase; $\omega_c \in \Omega$, the claimed identity; $t_c \in T$, the corresponding template; and ω, the real identity (that may not belong to Ω, in case of impostors). The goal is to decide whether the hypothesis $\omega_c = \omega$ is true or false. Due to the small amount of data usually available for each user (often, a single template), verification is done by matching x and t_c through a similarity measure. If the corresponding matching score $S(x, t_c)$ is lower than a threshold θ (estimated from training data), the user is rejected as an impostor, otherwise he/she is accepted as a genuine user [14].

Recently, it has been argued that biometric verification can be tackled more effectively as a set of N distinct two-class classification [4] or hypothesis testing [16] problems (one for each user model). Constructing a distinct classifier for each identity is however not straightforward, since only a small amount of data is available for each user. Further, for many biometric traits, only a similarity measure between samples (i.e., images) can be exploited to this end, as a representation in terms of feature vectors of fixed dimensionality, suitable for classification, may not be available (e.g., minutia-based fingerprint matching).

In this paper, we consider a possible approach to implement biometric verification as N two-class classification problems, which can be applied even if only similarities between samples are available, in the absence of a feature vector representation suitable for classification. It is based on the idea that a similarity measure $S(x', x'')$ between samples can be interpreted under some conditions as a kernel function $K(x', x'')$ that corresponds to the inner product between a mapping of x' and x'' into an implicitly defined feature space of fixed (though possibly infinite) dimensionality. This enables the exploitation of all the theory and methods in the field of kernel machines. In particular, in the case of biometric verification, this makes it possible to learn a two-class classifier, like a support vector machine (SVM), for each client, using the templates of the corresponding identity as positive examples, and the ones of the other identities as negative examples. Even if a given similarity measure can not be directly interpreted as a proper kernel function (i.e., a *definite* kernel), methods developed in the field of similarity-based classification can be exploited to obtain a proper kernel function, or to learn kernel machines with *indefinite* kernels [7, 9].

Interpreting a biometric similarity measure as a kernel function has many other potential advantages, besides the above described one: it does not require to define an application-specific specific kernel; the *sparsity* of the SVM solution allows the system to compute a small number of matching scores at verification phase, which is an appealing feature due to strict constraints on processing time; and it allows one to exploit incremental learning algorithms to efficiently update each user model (classifier), when needed.

It turns out that the above approach has also an interesting relationship with cohort-based biometric verification methods [1, 3, 18, 12], that exploit also matching scores with respect to other templates than t_c. A potential advantage of our approach is that the cohort size and elements are automatically selected in a principled way by the SVM learning algorithm, instead of heuristically. It also allows us to provide an intuitive motivation to existing heuristic criteria for cohort selection.

Although we formulate our approach in a verification setting, we show by experiments that it can also be exploited for biometric identification, where the goal is to identify the user identity based solely on the submitted biometric trait.

The paper is structured as follows. In Sect. 2 we overview cohort-based methods, due to their relationship with our approach, that is described in Sect. 3. Experimental results are reported in Sect. 4. Conclusions and future work are discussed in Sect. 5.

2. Cohort-based biometric verification

The traditional approach to biometric verification completely disregards information about the "background" class, namely, the so-called "non-matching" (or "cohort") scores $S(x, t_i)$, $t_i \in T \setminus \{t_c\}$. In recent work, it has been shown that even using a small subset of cohort templates $t_i \in \mathcal{C} \subset T \setminus \{t_c\}$ can significantly improve the performance of biometric verification systems [1, 3, 18, 12]. Note that the cohort set $\mathcal{C}$ has to include only very few templates for computational efficiency at verification phase, as only a small number of matchings can be usually computed. Cohort-based biometric verification amounts to considering a distinct two-class classification problem for each client, as argued in [4], since the templates $t_i \in \mathcal{C}$ are selected separately for each client.

In [1] the likelihood ratio between the genuine and impostor distribution (which provides the optimal decision rule, in the Neyman-Person sense) was approximated, for each identity ω_c, as:

$$\frac{p(x|\omega_c = \omega)}{p(x|\omega_c \neq \omega)} \approx \frac{S(x, t_c)}{\max_{t_i \in \mathcal{C}} S(x, t_i)}, \tag{1}$$

where $\mathcal{C}$ is a small cohort set (e.g., 5 to 20 templates) whose templates are selected offline, as the most similar ones to t_c.

The above approach can be also interpreted as a user-dependent normalization, similarly to [3], where the so-called Test-normalization (T-norm) was proposed. T-norm is similar to the widely known Z-normalization, but its parameters are computed during verification instead of being determined from training data (i.e., using only the enrolled templates). It is given by:

$$\frac{S(x, t_c) - \mu}{\sigma}, \tag{2}$$

where μ and σ are respectively the mean and standard deviation of the matching scores $S(x, t_i)$, $t_i \in \mathcal{C}$.

In [1, 18] a discriminative approach was proposed, instead, which can be viewed as the estimation of the following posterior with a classifier, whose features are the matching score $S(x, t_c)$ and the cohort scores:

$$P(\omega_c = \omega | S(x, t_c), \{S(x, t_i) : t_i \in \mathcal{C}\}). \tag{3}$$

Although experimental evidences proved that cohort-based methods can outperform the traditional approach, they are mainly based on heuristics, and on very restrictive assumptions. In particular, the selection of the cohort set is not based on any well-grounded prototype selection method, or optimality criterion (e.g., the maximization of classification accuracy). Recently, it has been experimentally shown that also cohorts that are not similar to the claimed identity can indeed exhibit some discriminant capability [12]. We will provide a first motivation to this experimental evidence, based on the geometric interpretation of SVMs as optimal separators of convex hulls [9].

3. Learning SVMs with biometric similarity

The main idea of our work is to interpret the similarity measure $S(x', x'')$ of a biometric matcher between two biometric traits $x', x'' \in \mathcal{X}$, as a kernel function $K(x', x'')$, which corresponds to the inner product between a mapping $\phi : \mathcal{X} \mapsto \Phi$ of x' and x'' into an embedded feature space Φ of fixed (though possibly infinite) dimensionality. This allows us to learn kernel machines for biometric verification and identification.

We approach the verification problem as follows. As suggested in [4], we learn a two-class classifier for each client. The goal of each classifier is to decide whether the corresponding identity, when claimed, is true (genuine claim) or not (impostor claim). Accordingly, the training set of each classifier consists of a set of genuine and impostor *samples* (e.g., face images) for the corresponding identity $\omega_c \in \Omega$, which can be obtained using the enrolled templates. If only one template per identity is available, the training set is made up of t_c as the only positive (genuine) sample, and $T \setminus \{t_c\}$ as $N - 1$ negative (impostor) ones.

A natural choice is to use SVMs as the base two-class classifier. Denoting with $y = +1$ and $y = -1$ respectively the "genuine" and "impostor" labels, the resulting SVM output for identity ω_c is given by:

$$g_c(x) = \sum_{i=1}^{N} y_i \alpha_i K(x, t_i) + b, \tag{4}$$

where $y_c = +1$, $y_j = -1$, $j \neq c$, the coefficients b and $\alpha_i \in [0, C]$ are set by the learning algorithm, and C is the regularization parameter that tunes the trade-off between the training error and the margin [19]. The above coefficients are obtained by solving the following convex optimization problem (dual formulation):

$$\begin{aligned} \min_{\alpha} \quad & \frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N} y_i y_j \alpha_i \alpha_j K(t_i, t_j) - \sum_{i=1}^{N} \alpha_i \\ \text{s. t.} \quad & \sum_{i=1}^{N} \alpha_i y_i = 0, \;\; 0 \le \alpha_i \le C, \; i = 1, \ldots, N. \end{aligned} \tag{5}$$

The decision function is given by sign $g_c(x)$ [19].

The solution to problem 5 is sparse, namely, it only depends on a small subset of training samples for which $\alpha_i > 0$, called support vectors (SVs). In particular, if $0 < \alpha_i < C$, then $y_i g_c(x_i) = 1$. This means that x_i lies exactly on the margin, and it is thus called *margin* support vector. If $\alpha_i = C$, then $y_i g_c(x_i) < 1$; x_i violates the margin, and it is potentially misclassified.[1] It is thus referred to as *error* vector. Note that problem 5 and the resulting discriminant function 4 require one to compute only inner products between samples in the feature space Φ through the kernel function $K(\cdot, \cdot)$, without even knowing the mapping function ϕ explicitly. This is known as the *kernel trick*.

In the verification setting, we learn N SVMs, one for each identity, and verify whether x belongs to the claimed identity (genuine) or not (impostor) by only considering the corresponding SVM. In the next section, we experimentally show that our approach can also be successfully exploited for biometric identification, where we aim to find the identity ω which x belongs to. However, as this is a very difficult problem, usually a candidate list of M identities ($1 \le M < N$) is returned by the system, instead of a single identity. In our case, x is thus evaluated by all the N classifiers, and the outputs $g_1(x), \ldots, g_N(x)$ are sorted in descending order. The first M values of this list are selected as our best candidate list. Since the outputs of different SVMs may have different scales (e.g., if different values of the regularization parameter C are used), posterior probability estimates can be considered, instead [13].

The choice of SVMs as the base classifiers is not only motivated by the fact that they have achieved very high classification accuracies in several applications, but also that they exhibit several interesting advantages suited to biometric applications. First, thank to the kernel trick, SVMs can be applied even if a feature vector representation of samples is not available, as usually happens in biometrics tasks. Second, the sparsity of the SVM solution implies that only a *small* number of kernel evaluations (matchings) have to be computed to classify a sample. This is an appealing feature for biometric verification, due to strict constraints on processing time. Moreover, if application requirements force us to compute an even smaller number of kernel evaluations, techniques for reducing the number of SVs can be exploited, without degrading significantly the classification accuracy [8, 17]. Third, once an SVM is learnt, its decision function can be efficiently updated through incremental learning to account for insertion of novel training samples [5]. This is very useful for biometric systems as well, when the template galleries of enrolled users are frequently updated. Finally, incremental learning can also be exploited for the estimation of the C parameter in a much more efficient way than using the standard cross-validation, as well as for training large-scale SVMs [10].[2]

Our approach has also some interesting connections with cohort-based verification methods. In particular, the SVs of the negative class in the SVM of each client can be clearly interpreted as the cohort set of that client. This means that, in our approach, the cohorts are selected in a principled way (i.e., by minimizing the SVM objective function), instead of heuristically, and that their "optimal" number is determined automatically, for each client. For instance, this allows the cohort sets of those clients whose identity is more difficult to verify to contain more templates, while existing methods consider a fixed number of cohorts for each client. Note that the number of cohorts, that is already small due to the sparsity of the SVM solution, can be further reduced by exploiting the above mentioned techniques for pruning the set of SVs. In the next section we also discuss some interesting insights on the selection of the cohort set, by looking at the geometrical interpretation of the SVM.

Finally, we point out that only functions $K : \mathcal{X} \times \mathcal{X} \mapsto \mathbb{R}$ that obey the Mercer theorem [19] are proper (definite) kernels. Equivalently, the $N \times N$ matrix $K(t_i, t_j)$, $t_i, t_j \in T$, has to be positive semi-definite (psd). If this is not the case, one can resort to the framework of similarity-based classification [7], that discusses a number of approximations, based on sound theoretical motivations, to build kernel functions from non-psd similarity measures, namely, from *indefinite* kernels. Indefinite kernels can be even used directly to learn SVMs [7, 9, 13]. A theoretical motivation was given in [9]. It was shown that learning SVMs with indefinite kernels amounts to minimizing the distance between the convex hulls of the two classes of training samples in a pseudo-Euclidean space, which is a generalization of the concept of margin maximization.

4. Experiments and discussion

In this section we report an experimental investigation of the proposed approach in biometric verification and identification, considering two data sets of fingerprint and face images.

[1] Note that a classification error only occurs when $y_i g_c(x_i) < 0$.

[2] Note that C is the only parameter to be set in our approach, as the kernel is implicitly defined by the biometric similarity measure, and, thus, no kernel parameters have to be set.

Fingerprint data set. This data set consists of 142 distinct clients (by "client", here, we mean a distinct finger, even if it belongs to the same person). For each finger, twenty different impressions were acquired in two different sessions, separated by about two weeks. Only four fingers were considered in this case: the left and right index and thumb fingers. The fingerprint images were acquired using the Biometrika FX2000 optical sensor. This data set was also used in the context of the First International Competition on Fingerprint Liveness Detection [11]. The NIST Bozorth3[3] minutia-based matching algorithm was used to compute the similarity between fingerprint images.

Face data set. We used part of the data described in [2]. In particular, we considered the 200 video clips of real-access attempts from 50 different clients, acquired under different lighting conditions (and disregarded the facial spoofing attacks). As we need to operate on images, we extracted 12 face images for each client from each video clip. The Elastic Bunch Graph Matching (EBGM) algorithm was used for face verification.[4]

The training sets were built in both cases by randomly selecting a subset of $m = 1, 2$, or 5 templates per identity from the available data, and using the remaining images for testing. The results were averaged over 5 repetitions, considering different pairs of training and testing sets.

We compared our approach with the baseline matching algorithm ('baseline'), and the cohort-based approaches described by Eqs. 1 ('aggarwal-max') and 2 ('tnorm') in Sect. 2. When more than one template per identity was available, the baseline method was obtained by computing the average of the matching scores between the test sample x and all the templates of the claimed identity. Under the same condition, the cohort set of each identity was chosen by selecting a number of templates from other identities, that were the closest to any of the templates of the considered identity. More precisely, if $T^{\mathrm{c}} = \{t_1^{\mathrm{c}}, \dots, t_m^{\mathrm{c}}\}$ is the set of templates of the considered identity, for each template $t^{\mathrm{i}} \notin T^{\mathrm{c}}$, an "aggregate" similarity was computed as $\max\left(S(t^{\mathrm{i}}, t_1^{\mathrm{c}}), \dots, S(t^{\mathrm{i}}, t_m^{\mathrm{c}})\right)$. Then, the templates t^{i} were sorted according to the descending order of aggregate similarity, and the first $|\mathcal{C}|$ were used as cohorts. Note that no well-grounded method for selecting cohorts in the case of multiple templates per identity has been proposed thus far. Hence, we tried to extend the notion of "closeness" to a template of the claimed identity, that is the underlying idea of cohort selection [1], to the case of multiple templates. Further, we also computed the aggregate similarity for selecting cohorts using the mean instead of the maximum, but this choice slightly worsened the performance of the considered cohort-based methods on our data sets. We omitted these results for the sake of space.

To implement our approach, we used the well-known LibSVM software [6]. We always used the biometric similarity measure at hand as the kernel function, as the algorithm of LibSVM converges also with indefinite kernels [13]. Further, performances obtained by clipping the spectrum of the similarity matrix were not significantly different. The regularization parameter C of the SVMs was set to very high values (e.g., 10^6) as suggested by cross-validation. Note that this choice also provides *sparser* solutions. Since the two classes were highly unbalanced (one identity vs all), the value of C was weighted differently for each class. In particular, for each class, C was multiplied by the prior probability of the opposite class, estimated from training data. This is quite common to balance the contribution of the two classes to the classification error.

In these experiments, we compared our approach (denoted as 'svm') also with two variants of it. One, denoted as 'svm-cohort', learns an SVM for each user using only the templates of that user and the corresponding cohorts, selected as explained above. This is useful to understand to what extent the use of SVMs may outperform 'aggarwal-max' and 'tnorm', exploiting the same templates. The other, denoted as 'svm-reduced', consists of reducing the number of SVs, and allows us to assess the performance degradation with respect to the unpruned case ('svm'). To this end, we used a similar approach to that proposed in [8], where the underlying idea was to retrain an SVM only with the *margin* SVs (see Sect. 3). However, in our case, this was not enough to reduce the number of SVs, as the majority of them were exactly margin SVs. Thus, we decided to retain all the SVs of the positive class (i.e., the identity to be verified), and only those which were assigned the highest α values in the negative class (i.e., the cohort set).

We fixed the number of cohorts to 5 for the face experiment, and 15 for the fingerprint case. Note that the total number of matchings was given by the number of templates per user, for the baseline matcher; and the number of templates per user plus the number of cohorts, for cohort-based methods, 'svm-cohort' and 'svm-reduced' (at maximum). The method 'svm' was trained on the complete training set, without restriction on the number of matchings, to show the maximum performance achievable by our approach. Nevertheless, since the number of matchings is not always required to be small, in some cases, the method 'svm' could be applied directly; e.g., in *offline* identification problems.

Results for biometric verification and identification are reported in Fig. 1 and 2, according to standard performance evaluation measures: detection error trade-off (DET) curves (False Rejection Rate, FRR, vs False Acceptance Rate, FAR) and cumulative matching characteristic (CMC) curves (probability of identification vs candidate list size). The DET curves were obtained by varying the decision

[3] http://www.nist.gov/itl/iad/ig/nbis.cfm

[4] http://www.cs.colostate.edu/evalfacerec/algorithms5.php

threshold. The CMC curves were obtained by sorting the classifiers' outputs, or normalized scores, in descending order, to generate the candidate list.

For both face and fingerprint biometrics, and in both verification and identification, the performance of the proposed methods ('svm' and 'svm-reduced') increased as the number of templates per identity increased. The performance of our methods clearly outperformed the other methods on the face data set; for instance, in the identification case with 5 templates per identity, the probability of correct identification was greater than 90% with a very small candidate list size. Conversely, on the fingerprint data set, the performance of 'svm-reduced' was significantly worse than the other approaches when only 1 or 2 templates per identity were used. The reason is that, in this case, the algorithm used to reduce the number of SVs was unable to produce a very compact set of SVs without degrading the accuracy. On the other hand, though with an unacceptable number of matchings, 'svm' always outperformed the other approaches, highlighting that there might be room for improvement.

Lastly, we note that 'svm-reduced' always outperformed 'svm-cohort'. Moreover, the number of matchings performed by 'svm-cohort' was very often slightly lower than 'svm-reduced'. This means that the cohort set selected heuristically included some useless (or redundant) templates, which did not join the set of SVs, and, thus, did not contribute to the decision function. In other words, the heuristic selection of the cohort set may lead to suboptimal choices. This can also be intuitively motivated by a recent geometrical interpretation of SVMs. The SVM algorithm has proven to provide an optimal separator between the (reduced) convex hulls that respectively enclose the positive and the negative samples [9]. The SVs of each class are chosen to form a subset of the vertices of the convex hull which they belong to. Accordingly, the SVs of one class may include some of the closest points to the opposite class, as well as some of the farthest points to it. This means that not only the closest cohorts to the templates of a given identity (in terms of matching score) can be exploited for improving biometric recognition, but also the farthest ones, which confirms the experimental findings in [12].

5. Conclusions and future work

We investigated the application of similarity-based classification to biometric verification and identification, exploiting similarity functions (which are almost always available in biometric tasks) as kernel functions. This naturally allowed us to implement these tasks as a distinct two-class classification problem for each user model, as recently suggested in [4]. The proposed approach was able to improve the performance of biometric verification and identification systems at the expense of a small number of "non-match" scores, comparable with the one of cohort-based methods in verification tasks. We also showed that cohort selection can be faced in a more principled way, exploiting the geometric properties of the SVM algorithm, and, in general, of kernel-based methods.

Our approach also opens several interesting research issues. For instance, the automatic selection of a compact set of representative samples (the SVs) for each user model may be exploited in the context of template selection. Moreover, principled guidelines for aggregating multiple templates (see, e.g., [15]) may be derived by exploiting methods that reduce the number of SVs through linear combination in the embedded feature space [17].

References

[1] G. Aggarwal, N. Ratha, and R. Bolle. Biometric verification: Looking beyond raw similarity scores. In *Comp. Vision and Patt. Rec. Workshop*, 2006. 2, 4

[2] A. Anjos and S. Marcel. Counter-measures to photo attacks in face recognition: A public database and a baseline. In *Int'l J. Conf. Biometrics*, pp. 1–7, 2011. 4

[3] R. Auckenthaler, M. Carey, and H. Lloyd-Thomas. Score normalization for text-independent speaker verification systems. *Dig. Sig. Proc.*, 10:42–54, 2000. 2

[4] S. Bengio and J. Mariéthoz. Biometric person authentication is a multiple classifier problem. In *7th Int'l Conf. Multiple Classifier Sys.*, LNCS, pp. 513–522, Springer-Verlag, 2007. 1, 2, 5

[5] G. Cauwenberghs and T. Poggio. Incremental and decremental support vector machine learning. In *NIPS*, pp. 409–415. MIT Press, 2000. 3

[6] C.-C. Chang and C.-J. Lin. LibSVM: a library for support vector machines, 2001. 4

[7] Y. Chen, E. K. Garcia, M. R. Gupta, A. Rahimi, and L. Cazzanti. Similarity-based classification: Concepts and algorithms. *J. Mach. Learn. Res.*, 10:747–776, 2009. 1, 3

[8] D. Geebelen, J. A. K. Suykens, and J. Vandewalle. Reducing the number of support vectors of SVM classifiers using the smoothed separable case approximation. *IEEE Trans. Neural Netw. Learning Syst.*, 23(4):682–688, 2012. 3, 4

[9] B. Haasdonk. Feature space interpretation of svms with indefinite kernels. *IEEE Trans. Patt. Analysis and Mach. Intell.*, 27(4):482 –492, 2005. 1, 2, 3, 5

[10] T. Hastie, S. Rosset, R. Tibshirani, and J. Zhu. The entire regularization path for the support vector machine. *J. Mach. Learn. Res.*, 5:1391–1415, 2004. 3

[11] G. L. Marcialis, A. Lewicke, B. Tan, P. Coli, D. Grimberg, A. Congiu, A. Tidu, F. Roli, and S. A. C. Schuck-

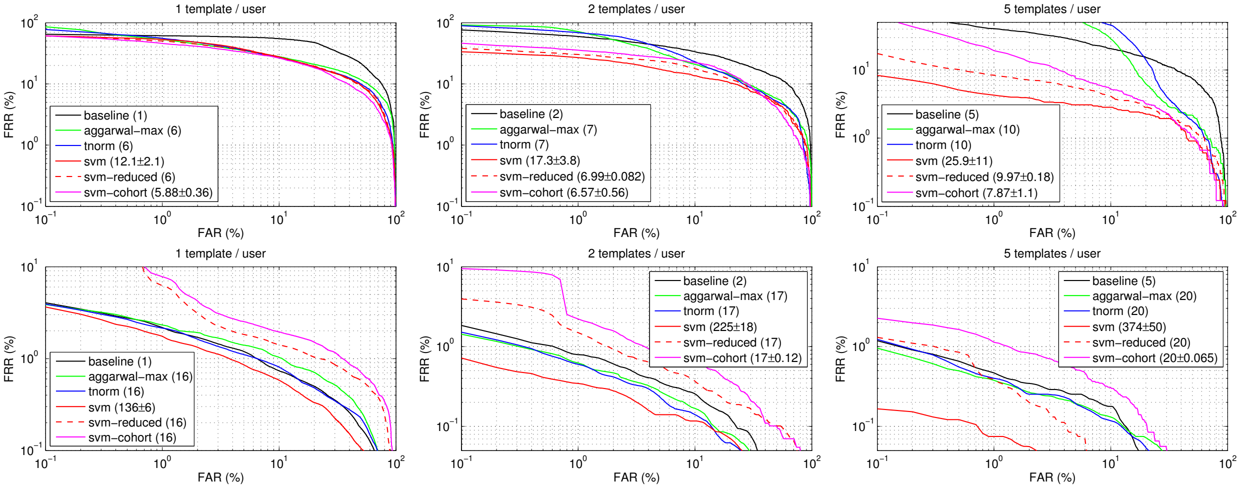

Figure 1. Results (DET curves) for face verification (top) and fingerprint verification (bottom). The number of matchings at test time is reported in the legend, as mean and standard deviation for SVM-based methods.

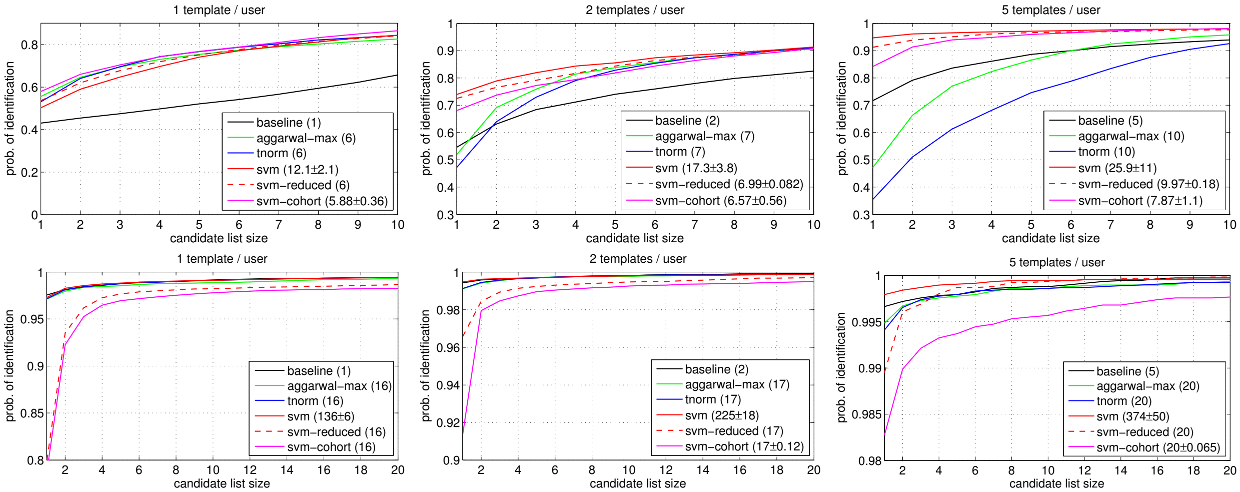

Figure 2. Results (CMC curves) for face identification (top) and fingerprint identification (bottom). The number of matchings at test time is reported in the legend, as mean and standard deviation for SVM-based methods.

ers. 1st Int'l Fingerprint Liveness Detection Comp. In *ICIAP*, vol. 5716 *LNCS*, pp. 12–23. Springer, 2009. 4

[12] A. Merati, N. Poh, and J. Kittler. Extracting discriminative information from cohort models. In *IEEE 4th Int'l Conf. Biom.: Theory, App. and Sys.*, 2010. 2, 5

[13] J. Platt. Probabilistic outputs for SVMs and comparison to regularized likelihood methods. In *Adv. Large Margin Classifiers*, pp. 61–74, 2000. 3, 4

[14] S. Prabhakar and A. K. Jain. Decision-level fusion in fingerprint verification. *P. Rec.*, pp. 861–874, 2002. 1

[15] C. Ryu, H. Kim, and A. K. Jain. Template adaptation based fingerprint verification. In *18th Int'l Conf. Patt. Rec.*, vol. 04, pp. 582–585, 2006. IEEE CS. 5

[16] W. J. Scheirer, A. Rocha, R. Michaels, and T. E. Boult. Meta-recognition: The theory and practice of recognition score analysis. *IEEE Trans. on Patt. Analysis and Mach. Intell.*, 33:1689–1695, 2011. 1

[17] B. Schölkopf, S. Mika, C. J. C. Burges, P. Knirsch, K.-R. Muller, G. Rätsch, and A. J. Smola. Input space versus feature space in kernel-based methods. *IEEE Trans. on Neural Netw.*, 10(5):1000–1017, 1999. 3, 5

[18] S. Tulyakov, Z. Zhang, and V. Govindaraju. Comparison of combination methods utilizing T-normalization and second best score model. In *CVPR*, 2008. 2

[19] V. N. Vapnik. *The nature of statistical learning theory*. Springer-Verlag, NY, USA, 1995. 3