Signature verification based on fusion of on-line and off-line kernels

Vadim Mottl, Mikhail Lange
Computing Center RAS, Moscow, Russia
vmottl@yandex.ru

Valentina Sulimova, Alexey Yermakov
Tula State University, Tula, Russia
vsulimova@yandex.ru

Abstract

The problem of signature verification is considered within the bounds of the kernel-based methodology of pattern recognition, more specifically, SVM principle of machine learning. A kernel in the set of signatures can be defined in different ways and it is impossible to choose the most appropriate kernel a priori. We propose a principle of fusing several on-line and off-line kernels into an entire training and verification technique. Experiments with signature database SVC2004 have shown that the multi-kernel approach essentially decreases the error rate in comparison with verification based on single kernels.

1. Introduction

The problem of signature verification consists in testing the hypothesis that a given signature belongs to the person having claimed his/her identity. Depending on the initial data representation, it is adopted to distinguish between on-line and off-line signature verification [1].

During more than 20 years long history of studying the problem of signature verification, a plenty of ideas have been proposed and tested, practically all of which fall into two groups: feature-based [2,3,4] and function-based [5,6] methods. However, any method of signature verification is based, finally, on a metric in the set of signatures.

As a rule, it is impossible to known in advance which of possible metrics is more appropriate for a concrete person. Therefore, in a number of papers it is proposed to combine several methods of signature verification [2,3,4,7,8].

In this paper, we apply a natural kernel-based way of easily combining on-line and off-line methods into an

entire verification technique. The main notion of this approach proposed in [9] is that of a kernel function defined as a symmetric two-argument function possessing the property to form positive semidefinite matrix for any finite collection of entities [10].

Any kernel function defined in the set of signatures embeds it into some hypothetical linear space in which it plays the role of inner product [11]. This property of the kernel function allows for reformulating practically any of existing methods of signature verification in kernel-based terms and to combine several different methods in the process of constructing a joint decision rule.

Such a way belongs to the group of so-called sensor-level techniques of modality fusion, which, in accordance with general investigations in the field of multimodal biometrics [12] and our previous experiments [13], can yield better results in comparison with fusion of modality-specific classifiers, in particular, on the level of classifier scores [3,4] or decision rules [7,8]. Besides, the kernel fusion technique we propose here allows to avoid the computationally hard problem of quadratically-constrained quadratic optimization which arises when alternative kernel fusion techniques are applied [14,15].

In the previous work [13], we demonstrated the advantages of the multi-kernel approach to the problem of on-line signature verification. In this paper, the kernel-based approach is extended onto the problem of combining the on-line and off-line modalities into an entire signature verification technique. Experiments with the data base of the World Signature Verification Competition (SVC2004) [16] show that the combined technique provided essential decreasing of the error rates achievable within the bounds of each single modality.

2. Metrics and kernels in the set of signatures

2.1. Metric in the set of on-line signatures

Each on-line signature is represented by a multi-component vector signal $\alpha = (\mathbf{x}_s, s = 1,...,N)$ which

initially includes five components: pen tip coordinates (X and Y), pen tilt azimuth (Az) and altitude (Alt), and pen pressure (Pr) (Figure 1). We supplement the signals with two additional variables – pen's velocity and acceleration.

Figure 1. Off-line (images) and on-line (signals) representation of signatures.

For comparing pairs of signals of different lengths $\left[\alpha'=(\mathbf{x}'_s,\ s=1,...,N'),\ \alpha''=(\mathbf{x}''_s, s=1,...,N'')\right]$, we use the principle of dynamic time warping [5] with the purpose of aligning the vector sequences.

Each version of alignment $w(\alpha',\alpha'')$ is equivalent to a renumbering the elements in both sequences $\alpha'_w=(\mathbf{x}'_{w,s'_k}, k=1,...,N_w)$, $\alpha''_w=(\mathbf{x}''_{w,s''_k},\ k=1,...,N_w)$, $N_w \geq N'$, $N_w \geq N''$. Let W be the set of all alignments of two signals α' and α''. The best alignment $\hat{w}(\alpha',\alpha'')$ is defined by the condition

$$\hat{w}(\alpha',\alpha'') = \arg\min_{w\in W}\left\{\sum_{k=1}^{N_w}\|\mathbf{x}'_{w,s'_k}-\mathbf{x}''_{w,s''_k}\|^2 + \beta\sum_{k=2}^{N_w}\left(I[s'_{k+1}=s'_k]+I[s''_{k+1}=s''_k]\right)\right\}, \quad (1)$$

where $I[\cdots]$ is indicator function which equals 1 if the condition in brackets is met and 0 otherwise and β is penalty upon each repetition of elements.

It is easy to prove that the function defined by the best alignment (1) as

$$\rho(\alpha',\alpha'') = \sqrt{\sum_{k=1}^{N_{\hat{w}}}\|\mathbf{x}'_{\hat{w},s'_k}-\mathbf{x}''_{\hat{w},s''_k}\|^2} \quad (2)$$

satisfies all the properties of a metric.

2.2. Metric in the set of off-line signatures

For comparing grayscale images (patterns) representing off-line signatures we, apply the technique of tree-structured pattern representation proposed in [6].

For the given pattern P, the recursive scheme described in [6] produces a pattern representation R in the form of a complete binary tree of elliptic primitives (nodes) Q: $R=\{Q_n : 0 \leq n \leq n_{\max}\}$, where n is the node number of the level $l_n = \lfloor \log_2(n+1)\rfloor$.

For comparing any two nodes Q'_n and Q''_n, a dissimilarity function $d(Q'_n,Q''_n) \geq 0$ can be easily

defined through parameters of these primitives such as center vectors and appropriate pairs of direction vectors [6]. Using it we define a loss function

$$D(Q'_n,Q''_n)=\begin{cases} d(Q'_n,Q''_n), \text{ if } Q'_n \text{ and/or } Q''_n \text{ are "end" nodes,} \\ 0, \text{ otherwise.}\end{cases}$$

Then, following [6], we define the dissimilarity measure (metric) of the trees R' and R'' as

$$\rho(R',R'') = \sum_{R'\cap R''} 2^{-l_n} D(Q'_n,Q''_n), \quad (3)$$

where the sum is taken over all pairs $(Q'_n,Q''_n)\in(R'\cap R'')$.

2.3. Transformation of a metric into kernel

Let $\omega',\omega'' \in \Omega$ be two signatures represented by signals (α',α'') or trees (R',R''). Whereas metric (2) or (3) evaluates some dissimilarity of signatures, function

$$K(\omega',\omega'') = \exp\left[-\gamma\rho^2(\omega',\omega'')\right] \quad (4)$$

will have the sense of their pair-wise similarity. If coefficient γ is large enough, this function for any finite collection of signatures will form positive semidefinite matrix $[K(\omega_i,\omega_j);\ i,j=1,...,N]$, i.e. it is a kernel function in the set of signatures Ω. Function (4) is usually called the radial kernel function.

The notion of a kernel $K(\omega',\omega''),\ \omega',\omega'' \in \Omega$ in the set of signatures $\omega \in \Omega$ allows for treating it a subset in some real linear space $\Omega \subseteq \tilde{\Omega}$ in which it plays the role of inner product.

2.4. The kernels studied in experiments

It should be noticed that formula (4) implies actually a family of kernels because different metrics may occur in it.

In our experiments concerned with the problem of signature verification, we studied 13 radial kernels (Table 1). One of them is off-line radial kernel based on the metric (3). The others are on-line kernels which differ from each other by the subset of utilized components of signals and by the value of the matching penalty β in the metric (2). The value of parameter γ was chosen identical $\gamma = 0.25$ in all the kernels.

Table 1. The kernels studied in the experiments

Kernel		Subset of components
$\beta = 10$	$\beta = 20$	
K_1	K_2	pen coordinates
K_3	K_4	pen tilt (azimuth and altitude)
K_5	K_6	pen pressure
K_7	K_8	coordinates, velocity, acceleration
K_9	K_{10}	coordinates, tilt, pressure
K_{11}	K_{12}	all seven components
Off-line kernel K_{13}		–

(The left side of the table rows K_1 through K_{12} are labeled "On-line kernels")

3. SVM for training in the linear space of signatures produced by a single kernel

In kernel terms, the SVM decision function [10] for classification of signatures into genuine ones $g=1$ and forgeries $g=-1$ can be represented as discriminant hyperlane

$$y(\omega)=K(\vartheta,\omega)+b>0\rightarrow g=1,\ \ y(\omega)\le 0\rightarrow g=-1, (5)$$

formed by the direction vector ϑ which is element of a hypothetical linear space $\vartheta\in\tilde{\Omega}$ into which the kernel $K(\omega',\omega'')$ embeds the set of signatures $\tilde{\Omega}\supset\Omega$. It can be found as linear combination of elements of the training set $\{\omega_j, j=1,...,N\}$:

$$\vartheta=\sum_{j:\lambda_j>0}g_j\lambda_j\omega_j$$

with coefficients $\lambda_j\ge 0$, which are solutions of the dual formulation of the SVM training problem

$$\begin{cases}\sum_{j=1}^{N}\lambda_j-(1/2)\sum_{j=1}^{N}\sum_{l=1}^{N}\left[g_jg_lK(\omega_j,\omega_l)\right]\lambda_j\lambda_l\rightarrow\max,\\ \sum_{j=1}^{N}g_j\lambda_j=0,\ 0\le\lambda_j\le C/2, j=1,...,N.\end{cases} (6)$$

4. Subset of relevance kernels resulting from kernel fusion

In this paper, we apply the kernel fusion technique proposed in [9]. This approach allows for choosing the most appropriate (relevant) kernels by analogy with constructing the Relevance Vector Machines (RVM) [17] which chooses the most relevant entities (vectors) in the training set.

Let $K_i(\omega',\omega'')$, $i=1,...,n$, be several kernel functions defined on the same set of signatures $\omega\in\Omega$. These kernels embed the set Ω into different linear spaces $\Omega\subset\tilde{\Omega}_i$, $i=1,...,n$. It is convenient to treat their jointly as Cartesian product

$$\tilde{\tilde{\Omega}}=\tilde{\Omega}_1\times...\times\tilde{\Omega}_n=\left\{\overline{\omega}=<\omega_1,...,\omega_n>:\omega_i\in\tilde{\Omega}_i\right\} (7)$$

formed by ordered n-tuples of elements from $\tilde{\Omega}_1,...,\tilde{\Omega}_n$.

The idea of adaptive training in the combined space [9] consists in jointly inferring the direction elements ϑ_i in the particular linear spaces $\tilde{\Omega}_i$ and the nonnegative weights r_i of the respective kernels by additionally penalizing large weights:

$$\begin{cases}\sum_{i=1}^{n}\left[(1/r_i)K_i(\vartheta_i,\vartheta_i)+\log r_i\right]+\\ C\sum_{j=1}^{N}\delta_j\rightarrow\min(\vartheta_1,...,\vartheta_n,r_1,...,r_n,b,\delta_j,j=1,...N), (8)\\ g_j\left[\sum_{i=1}^{n}K_i(\vartheta_i,\omega_j)+b\right]\ge 1-\delta_j,\delta_j\ge 0, j=1,...N.\end{cases}$$

This criterion displays a pronounced tendency to emphasize the kernels which are "adequate" to the training data and to suppress up to negligibly small values the weights r_i at "redundant" ones.

It can be shown [9] that the following iterative procedure solves the problem (8):

$$\vartheta_i^k=r_i^{k-1}\sum_{j:\lambda_j^k>0}g_j\lambda_j^k\omega_j,$$

$$r_i^k=(r_i^{k-1})^2\sum_{j:\lambda_j^k>0}\sum_{l:\lambda_l^k>0}K_i(\omega_j,\omega_l)\lambda_j^k\lambda_l^k.$$

At each iteration k, the coefficients $\lambda_1^k\ge 0,...,\ \lambda_N^k\ge 0$ are to be found as the solutions of the dual SVM problem having the structure analogous to (6):

$$\begin{cases}\sum_{j=1}^{N}\lambda_j-\frac{1}{2}\sum_{j=1}^{N}\sum_{l=1}^{N}\left[g_jg_l\sum_{i=1}^{n}r_i^kK(\omega_j,\omega_l)\right]\lambda_j\lambda_l\rightarrow\max,\\ \sum_{j=1}^{N}g_j\lambda_j=0,\ 0\le\lambda_j\le C/2, j=1,...,N.\end{cases}$$

Updating the constant b^k does not offer any difficulty. As a rule, the process converges in 10-15 steps.

5. Structure of experiments

In the experiment, we used the database of the Signature Verification Competition 2004 [16] that contains vector signals of 40 persons (Figure 1). On the basis of these signals we generated grayscale images (256×256 pixels) with 256 levels of brightness corresponding to the levels of pen pressure in the original signals.

For each person, the training set consists of 400 signatures, namely, 5 signatures of the respective person, 5 skilled forgeries, and 390 random forgeries formed by 195 original signatures of other 39 persons and 195 skilled forgeries for them. The test set for each person consists of 69 signatures, namely, 15 genuine signatures, 15 skilled forgeries, and 39 random forgeries. Thus, the total number of the test signatures for 40 persons amounts to 2760.

Twelve different on-line metrics and one off-line metric were simultaneously computed for each pair of signature signals (Section 2) and, respectively, thirteen different kernels were evaluated (Table 1).

We don't pursue here the aim to choose the "most appropriate" kernel providing the best accuracy of signature verification. Our aim is to show advantages of the approach utilizing several kernels at once, as against that based on a single predefined kernel.

6. Experimental results

We tested 14 ways of training, namely, based on each of the initial kernels $K_1(\omega',\omega''),...,K_{13}(\omega',\omega'')$ separately (Section 3) and with fusion of all the kernels (Section 4). The error rates in the total test set of 2760 signatures are shown in Table 2.

It is well seen that the combined kernel obtained by kernel fusion essentially outperforms each of the single ones. At the same time, for each of 40 persons whose signatures made the data set, the kernel fusion procedure has selected only one relevant kernel which turned out to be most adequate to his/her handwriting.

So, for each person the training procedure made the individual choice between the of-line and on-line modalities. For 12 persons the pictorial off-line signature represented by kernel K_{13} was recognized as more reliable, and in 28 cases one of the on-line kernels K_1 - K_{12} was preferred.

Figure 1 illustrates an obvious example of the situation when the on-line kernel K_6 based on pen-pressure information (Pr) is relevant for a specific person. It is well seen that two genuine signatures and two skilled forgeries have very similar off-line representations as well as four of five on-line signal components, and only the pen pressure dynamics clearly reveals the forgeries. The result of kernel fusion for this person is just kernel K_6 with the individual verification error 0%.

Table 2. Error rates for single kernels versus kernel fusion

Kernel	Errors %	Relevant as result of fusion for:	Kernel	Errors %	Relevant as result of fusion for:
K_1	0.65	5 persons	K_2	1.01	9 persons
K_3	5.58	0 persons	K_4	7.50	0 persons
K_5	2.75	0 persons	K_6	2.50	4 persons
K_7	0.98	2 persons	K_8	1.41	1 persons
K_9	0.36	4 persons	K_{10}	0.76	2 persons
K_{11}	0.47	1 persons	K_{12}	1.01	0 persons
K_{13}	1.63	12 persons	FUSION	**0.29**	-

7. Conclusions

The kernel-based approach to signature verification enables harnessing mathematically most advanced method of pattern recognition such as kernel-selective SVM. This approach predefines the algorithms of both training and recognition, and it remains only to choose the kernel produced by an appropriate metric in the set of signatures, such that the genuine signatures of the same person would be much closer to each other than those of different persons. However, different understandings of signature similarity lead to different kernels. The proposed kernel fusion technique automatically chooses the most appropriate kernel for each person in the process of adaptive training.

Experiments with signature data base SVC2004 demonstrate that verification results obtained by fusion of several on-line and off-line kernels in accordance with the proposed approach essentially outperforms the results based on single kernels.

References

[1] R. Plamondon, S. N. Srihari On-line and off-line handwriting recognition: A comprehensive survey. *IEEE Trans. on Pattern Recognition and Machine Intelligence, 22*(1), 2000, 63-84.

[2] J. Richiardi, H. Ketabdar, A. Drygajlo. Local and global feature selection for on-line signature verification. *Proceedings of the Eighth International Conference on Document Analysis and Recognition (ICDAR'05)*, Seoul, South Korea, 2005, 625-629.

[3] M. Fuentes, S. Garcia-Salicetti, B. Dorizzi. On line signature verification: fusion of a hidden Markov model and a neural network via a support vector machine. *Proceedings of the Eighth International Workshop on Frontiers in Handwriting Recognition (IWFHR'02).*

[4] L. Hu, Y. Wang. On-line signature verification based on fusion of global and local information. *Proceedings of the 2007 International Conference on Wavelet Analysis and Pattern Recognition*, 2007, 1192-1196.

[5] R.Martens, L. Claesen. Dynamic programming optimisation for on-line signature verification. *Proceedings of the Fourth International Conference on Document Analysis and Recognition (ICDAR'97), Ulm, Germany*, 1997, 2, 653-656.

[6] M. Lange, S. Ganebnykh, A.Lange. Moment-Based Pattern Representation using shape and grayscale features. *Lecture Notes in Computer Science*, Vol. 4477, Springer, 2007, pp. 523-530.

[7] A. Zimmer, L. Ling. A hybrid On/Off line handwritten signature verification system. *Proceedings of the 7-th International IEEE Conference on Document Analysis and Recognition (ICDAR'03), 2003.*

[8] I.Nakanishi, Y. Itoh, Y.Fukui. Multi-matcher on-line signature verification system in DWT domain. *ICASSP* 2005 *IEEE*, 965-968.

[9] Mottl V., Tatarchuk A., Sulimova V., Krasotkina O., Seredin O. Combining pattern recognition modalities at the sensor level via kernel fusion. *Proceedings of the 7th International Workshop on Multiple Classifier Systems.* Czech Academy of Sciences, Prague, Czech Republic, May 23-25, 2007.

[10] V. Vapnik. *Statistical Learning Theory* (John-Wiley & Sons, Inc., 1998)

[11] V. Mottl. Metric spaces admitting linear operations and inner product. *Doklady Mathematics*, 2003, 140–143.

[12] A. Ross, A.K. Jain. Multimodal biometrics: An overview. *Proceedings of the 12th European Signal Processing Conference (EUSIPCO),* Vienna, Austria, 2004, 1221-1224.

[13] V. Sulimova, V.Mottl, A.Tatarchuk. Multi-kernel approach to on-line signature verification. *Proceedings of the Eighth IASTED International Conference on Signal and Image Processing*, 2006, pp. 448-453

[14] G.R.G. Lanckriet, Cristianini N., Ghaoui L.E., Bartlett P., Jordan M.I. Learning the kernel matrix with semidefinite programming. *J. Machine Learning Research*, 5, 2004, 27–72.

[15] F.R. Bach, G.R.G. Lankriet, M.I. Jordan. Multiple kernel learning, conic duality, and the SMO algorithm. *Proceedings of the 21th International Conference on Machine Learning*, Banff, Canada, 2004.

[16] SVC 2004: First International Signature Verification Competition. *http://www.cs.ust.hk/svc2004/index.html*

[17] C.M. Bishop, M.E. Tipping. Variational relevance vector machines. *Proceedings of the 16th Conference on Uncertainty in Artificial Intelligence,* 2000, 46–53.