Face Authentication Using One-Class Support Vector Machines

Manuele Bicego[1,⋆], Enrico Grosso[1], and Massimo Tistarelli[2]

[1] DEIR, University of Sassari, via Torre Tonda 34, 07100 Sassari, Italy
Phone +39 079 2017321
bicego@uniss.it

[2] DAP, University of Sassari, piazza Duomo 6, 07041 Alghero (SS), Italy

Abstract. This paper proposes a new method for personal identity verification based the analysis of face images applying One Class Support Vector Machines. This is a recently introduced kernel method to build a unary classifier to be trained by using only positive examples, avoiding the sensible choice of the impostor set typical of standard binary Support Vector Machines. The features of this classifier and the application to face-based identity verification are described and an implementation presented. Several experiments have been performed on both standard and proprietary databases. The tests performed, also in comparison with a standard classifier built on Support Vector Machines, clearly show the potential of the proposed approach.

1 Introduction

Face analysis is undoubtedly an interesting research area for challenging methodological issues and potential practical applications. In the past, several methods have been proposed for face authentication (personal identity verification from one face sample) and recognition (subject identification from a face database), ranging from geometrical approaches to statistical or local analysis [1]: almost all techniques explored by the Pattern Recognition and Machine Learning research have been tested in the face analysis application context. Among others, Support Vector Machines [2] have had a great success in this field, for several reasons: fast training [3], accurate classification and, at the same time, high generalization capability (i.e. the ability of learning the trend and the regularity of observed data).

Support Vector Machines have been employed for face analysis, recognition [4, 5] and authentication [6, 7], showing promising results. Support Vector Machines are intrinsically a binary classifier: the SVM determines an hyperplane that optimally separates two sets, i.e. positive and negative examples. This is a limitation for identity verification because only positive examples should be used for training. The most widely adopted solution is to train a binary SVM with instances of the target face as positive examples and gathering a set of different faces from another database, as negative examples (representing "The rest of the

[⋆] Contact author

S.Z. Li et al. (Eds.): IWBRS 2005, LNCS 3781, pp. 15–22, 2005.
© Springer-Verlag Berlin Heidelberg 2005

world"). Consequently, the set of negative samples must be general enough to faithfully represent any individual but the target subject. Since the choice of a proper or complete set of negative examples is crucial, an improper or incomplete set could heavily bias the authentication result. In the experimental section of the paper, it will be shown that different choices of the negative set could lead to different authentication rates. Therefore, a desirable solution is represented by a modelling tool, which maintains the good classification capabilities of the SVM, but also to be trained by positive examples only. Such tool exists, and is represented by the recently introduced One Class Support Vector Machines [9–11]: this method determines the optimal hypersphere enclosing the training data in the feature space. An interesting application to face detection has been proposed by Jin and others [12]

This paper investigates the use of one class SVM for identity verification from face images. Image feature vectors are used to train a One-Class SVM. In the authentication phase, the face vector is fed to the trained classifier and the distance from the center of the hypersphere is computed. This measure represents the inverse of the matching score. The claimed identity is verified by comparing the score with a pre-determined threshold. In the experimental section, OC-SVM and standard binary SVM are compared using the ORL database. Further experiments on the OC-SVM based system have been also reported.

2 One Class Support Vector Machines

One Class SVM (OC-SVM) represents a recently introduced kernel method [9–11] able to characterize the support of a high dimensional distribution. The standard SVM is a binary classifier, able to distinguish between two classes by using a separating hyperplane. Due to its binary nature, this technique is not adequate to model one class problems, i.e problems in which only positive examples are present. On the other side, the OC-SVM is a unary classifier, able to derive a support vector description of a data set formed by positive examples only. In order to do that, OC-SVM computes the smallest hypersphere in the feature space enclosing the data. Also in this case, all the advantages of the kernel trick could be mantained.

More formally, let $D = \{x_i\}$ be the data set of ℓ examples to be modelled. In the linear case (no kernel is employed), the goal is to find the smallest sphere that contains the data, i.e:

$$||x_i - a||^2 \leq R^2 \qquad i = 1 \cdots \ell \tag{1}$$

where R is the radius, a is the center and $|| \cdot ||$ is the Euclidean norm.

These constraints can be relaxed by introducing the slack variables $\xi_i > 0$:

$$||x_i - a||^2 \leq R^2 + \xi_i \qquad i = 1 \cdots \ell \tag{2}$$

The problem of finding the smallest hyperplane is solved by introducing the Lagrangian $\mathcal{L}$:

$$\mathcal{L}(R, a, \xi_1, ..., \xi_\ell) = R^2 - \sum_i (R^2 + \xi_i - ||x_i - a||^2)\alpha_i - \sum_i \xi_i \beta_i + C \sum_i \xi_i \tag{3}$$

where $\alpha_i, \beta_i \geq 0 \; \forall i$ are the lagrange multipliers associated to the constrain (2), C is a trade-off parameter, and $\sum_i \xi_i$ is a penalty term accounting for the presence of the outliers. Setting to zero the partial derivative of $\mathcal{L}$ with respect to R, a, ξ_i and solving, we obtain

$$\sum_i \alpha_i = 1 \qquad a = \sum_i \alpha_i x_i \qquad \alpha_i = C - \beta_i \qquad \forall i \tag{4}$$

Using these results, the costrained minimization of the Lagrangian in (3) can be rewritten as the maximization of the Wolfe dual form $\mathcal{L}$:

$$\mathcal{L} = \sum_i \alpha_i x_i \cdot x_i - \sum_i \sum_j \alpha_i \alpha_j x_i x_j \tag{5}$$

$$0 \leq \alpha_i \leq C, \qquad \sum_i \alpha_i = 1 \qquad \forall i \tag{6}$$

Finally, the Karush-Kuhn-Tucker conditions give

$$\xi_i \beta_i = 0 \qquad (R^2 + \xi_i - ||x_i - a||^2)\alpha_i = 0 \tag{7}$$

The simple linear case could be quite easily extended to the non linear case. By looking to the dual form in (5) one could observe that the inner product between input points could be replaced by a Kernel function $K(x, x)$, exactly as in the standard SVM theory. The idea is to find the closing hypersphere in the space induced by the chosen kernel, resulting in a non linear boundary in the original space.

The only problem of this non linear case is that, if the mapping induced by the kernel is unknown (as in the Gaussian case), the center of the hypersphere could not be explicitly computed as a linear combination of the image of the training points. Nevertheless, the distance between a point x and the center of the hypersphere A can be computed in the feature point as:

$$d^2(x) = K(x, x) - 2\sum_i \alpha_i K(x_i, x) + \sum_i \sum_j \alpha_i \alpha_j K(x_i, x_j) \tag{8}$$

3 The Proposed Approach

Every authentication system is composed of two distinct modules: off-line enrollment and on-line verification. Both modules rely on the extraction of one feature vector from each image, as from the following steps:

1. The Viola face detector [13] is applied to detect face patterns on the image. It is based on a cascade of weak classifiers combined with the AdaBoost scheme, working in a multiscale fashion. If the face detector is unable to properly localize the face, the image is discarded from the subsequent analysis.
2. The gray level face image is normalized against global illumination changes.

3. The normalized face image is subsampled to obtain a fixed size image ($60 \times 60 pixels$). This procedure allows an approximate scale invariance.
4. The image is vectorized to obtain a feature vector. The direct use of raw gray level information as features for the Support Vector Machines is not new in the literature [14]).

One OC-SVM is trained for each subject, representing the subject template. A video stream is used for gathering all the training images and 50 images are automatically chosen for training. The choice may be performed randomly or driven by a knowledge-based criterion to choose the *most representative set of images.*

Also for the authentication phase several face images are acquired from a video sequence. Subsequently, all images are processed to obtain the feature vectors, which are fed to the OC-SVM corresponding to the subject of the claimed identity. Each image submitted to the OC-SVM produces a matching score, which is the distance from the center of the trained hypersphere. If this distance is below a given threshold the claimed subject's identity is confirmed. The choice of the authentication threshold is obviously critical. Two possibilities are considered: either to choose one threshold for all subjects or to set a different threshold for each subject. Both these options have been tested in the experimental section.

4 Experimental Results

This section reports a preliminary experimental evaluation of the proposed approach. In particular, in the first test the OC-SVM tool has been compared with standard SVM methods, using the ORL database. Subsequently, the system has been tested on two proprietary databases with varying facial expression, illumination and scale.

4.1 Binary Versus One-Class SVM

In order to perform a direct comparison between OC-SVM and standard binary SVM, the same data set, protocol and features are used to perform both tests. The ORL database, containing 40 subjects, with 10 images for each subject, has been used. For each subject only 5 images have been used for training. Both complete client and impostor tests were performed. In the first case the authenticator is tested using images of the same subject (but not the same used for training) to estimate the false rejection rate (FRR). In the impostor test, other subjects images are fed into the system, to measure the false acceptance rate (FAR). The equal error rate (EER), i.e. the rate for which FAR=FRR, is picked as an estimate of the discrimination capability of the classifier.

In standard applications of binary SVM the authentication phase is performed by computing the distance of the test image from the separating hyperplane. The sign of the distance determines the class to which the subject belongs. In the current implementation the distance from the hyperplane is used as matching measure, with the option of setting a threshold (which is 0 in standard SVM).

In binary SVM there is the need of defining the impostor set, i.e. the negative examples. In order to highlight the fact that the choice of the impostor set is crucial, four different training sessions were performed, varying the impostor set (for all 10 randomly chosen subjects, with 2 images each): from the Bern database[1], the Stirling database[2], the Yale database[3], and from a mixture of the Stirling and the Yale database.

Please note that all SVM parameters remained unchanged but only the impostor set was changed. In particular, for both SVM and OC-SVM, the Radial Basis Function kernel was used with σ and C parameters fixed. From the results displayed in Table 1 it could be noted that the choice of the impostor set, for binary SVM, is crucial. In fact, the classification results change when varying the impostor data set.

Table 1. Comparative EERs

Method	**Impostor Set**	**EER**
Binary SVM	Bern	6.00%
Binary SVM	Stirling	5.50%
Binary SVM	Yale	6.00%
Binary SVM	Mixed	6.19%
One class SVM	-	5.50%

The OC-SVM, instead, does not require any impostor set, and the performances are stable. In particular, the OC-SVM Equal Error Rate is equal to the best SVM case: OC-SVM maintains all the good classification capabilities of the SVM, still removing the awkward problem of needing an appropriate impostor set.

4.2 Proprietary Databases

The proposed system was thoroughly tested using two proprietary databases. The first database includes 25 subjects with an image sequence of 95 to 195 color images for each subject, with several changes in facial expression and scale (see fig. 1). The sampled images are 640x480 pixels. For the face classification the images have been reduced to gray level with 8 bits per pixel.

50 images have been chosen for training. The RBF kernel was used ($\sigma = 1800$), and the regularization constant C was set to 0.3. These parameters have been selected after a preliminary evaluation on the ORL database, varying the kernel and parameters' configurations.

[1] Downloadable from ftp://iamftp.unibe.ch/pub/Images/FaceImages

[2] Downloadable from http://pics.psych.stir.ac.uk

[3] Downloadable from http://pics.psych.stir.ac.uk

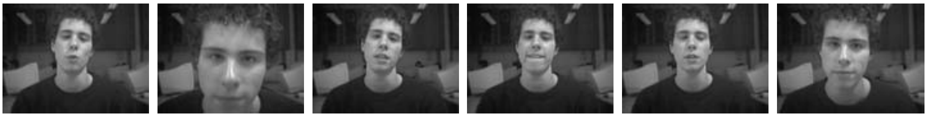
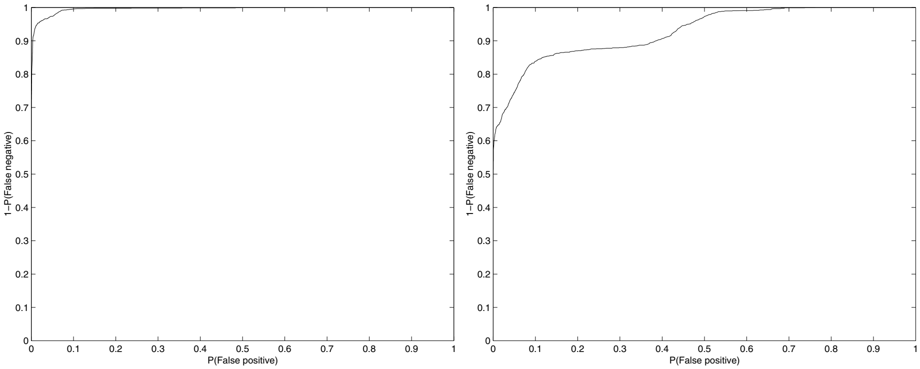


Fig. 1. Some variability in the first proprietary database

In table 2 the EER obtained for two different configurations is shown, i.e. by setting the same threshold for all subjects, or setting different thresholds for each subject. In the latter case, the reported EER is the average between all EERs resulting from the ROC curves. In fig. 2(a) the ROC curve for the single threshold selection is presented.

Table 2. EER from the proposed system based on One Class SVM, applied to the first proprietary database

Method	**number of client tests**	**number of impostor tests**	**EER**
One threshold for all subj	2275	122520	3.38%
One threshold for each subj	2275	122520	0.56%



(a) (b)

Fig. 2. ROC curves obtained from the identity verification test using: (a) the first database and (b) the second database

In order to better understand the potential of the proposed method a further experiment has been performed. The face image database has been augmented by adding more images to the existing set. For five out of the total 25 subjects, an image set acquired in a different session (at least two weeks apart) has been added. The training has been performed using only images from one of the two acquisition sessions. Some example images are presented in Fig. 3. The corresponding EERs are shown in table 3 for both single and multiple threshold selection. In fig. 2(b) the corresponding ROC curve for the single threshold selection is presented.

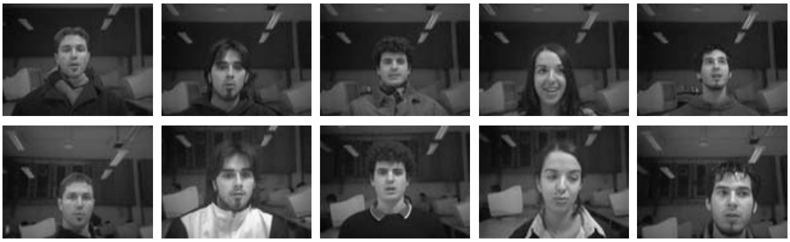


Fig. 3. Some example images from the two different sessions of the second proprietary database. Pictures from the same acquisition session are presented on the top and bottom row respectively

Table 3. EER from the proposed system based on One Class SVM, applied to the second proprietary database

Method	**number of client tests**	**number of impostor tests**	**EER**
One threshold for all subj	1760	10684	14.31%
One threshold for each	1760	10684	6.53%

5 Conclusions

In this paper a new face authentication technique has been proposed, based on One Class Support Vector Machines. This tool maintains all the attractive features of Kernel tools like standard Support Vector Machines, such as high generalization capabilities, but without the need for an impostor set. In the experimental evaluations it has been shown that, in the standard binary SVM case, the choice of the impostor set is crucial and may lead to different authentication performances.

A further improvement of the system is to combine, using a majority vote scheme, the results obtained by processing a whole video sequence. In other words, each image of a video sequence is processed, resulting in a yes/no answer. If the majority of the images produce a positive answer then the subject is authenticated, otherwise the authentication is rejected.

References

1. Zhao, W., Chellappa, R., Phillips, P., Rosenfeld, A.: Face recognition: A literature survey. ACM Computing Surveys **35** (2003) 399–458
2. Vapnik, V.: The Nature of Statistical Learning Theory. Springer-Verlag (1995)
3. Platt, J.: Fast training of support vector machines using sequential minimal optimization. In: Advances in Kernel Methods - Support Vector Learning. MIT Press (1999) 185–208
4. Guo, G., Li, S.Z., Kapluk, C.: Face recognition by support vector machines. Image and Vision Computing **19** (2001) 631–638

5. Heisele, B., Ho, P., Poggio, T.: Face recognition with support vector machines: Global versus component-based approach. In: Proc. Of IEEE Int. Conf. on Computer Vision. Volume 2. (2001) 688–694
6. Smeraldi, F., Capdevielle, N., Bigun, J.: Face authentication by retinotopic sampling of the gabor decomposition and support vector machines. In: Proc. of Int. Conf. on Audio and Video Based Biometric Person Authentication. (1999) 125–129
7. Jonsson, K., Kittler, J., Li, Y., Matas, J.: Support vector machines for face authentication. Image Vision Computing **20** (2002) 269–275
8. Tefas, A., Kotropoulos, C., Pitas, I.: Using support vector machines for face authentication based on elastic graph matching. In: Proc. of IEEE Int. Conf. on Image Processing. Volume 1. (2000) 29–32
9. Ben-Hur, A., Horn, D., Siegelmann, H., , Vapnik, V.: Support vector clustering. Journal of Machine Learning Research **2** (2001) 125–137
10. Schölkopf, B., Williamson, R., Smola, A., Shawe-Taylor, J., Platt, J.: Support vector method for novelty detection. In: Advances in Neural Information Processing Systems. Volume 12. (1999) 526–532
11. Tax, D., Duin, R.: Support vector domain description. Pattern Recognition Letters **20** (1999) 1191–1199
12. Jin, H., Liu, Q., Lu, H.: Face Detection Using One-Class-Based Support Vectors. In: Proc. of Sixth IEEE Int. Conf. on Automatic Face and Gesture Recognition (2004) 457–463
13. Viola, P., Jones, M.: Rapid object detection using a boosted cascade of simple features. In: IEEE Proc. Int. Conf. on Computer Vision and Pattern Recognition. Volume 1. (2001) 511–518
14. Pontil, M., Verri, A.: Support vector machines for 3d object recognition. IEEE Trans. on Pattern Analysis and Machine Intelligence **20** (1998) 637–646