EFFICIENT SCARF DETECTION PRIOR TO FACE RECOGNITION

Rui Min, Angela D'Angelo, and Jean-Luc Dugelay

EURECOM
Multimedia Communications Dept.
2229 route des Cretes BP 193
06904 Sophia Antipolis Cedex, FRANCE

ABSTRACT

Face occlusion is a very challenging problem in face recognition. The performance of face recognition system can decrease drastically due to the presence of partial occlusion on the face. One approach to overcome this problem is to first pre-classify faces into two classes: the clean face and the occluded face; then faces in different classes are treated by different recognition systems. In this case an algorithm which is able to automatically detect the presence of occlusions on the face will be a useful tool to increase the performances of the system. In this paper we present a scarf detection algorithm. In the experimental results the performances of the algorithm are reported and compared with state of the art systems.

1. INTRODUCTION

Over the past decades, face recognition has become a popular research topic in computer vision and one of the most successful applications of image analysis and understanding. Many researchers work actively and many face recognition methods have been proposed in the scientific literature [1].

State of the art face recognition systems perform with high accuracy under restricted environments, but performances drastically decrease in practical conditions such as video surveillance of crowded environments or large camera networks. The main problems are due to changes in facial expressions, illumination conditions, face pose variations and presence of occlusions on the selected face. Focusing on the last point, faces recorded through a video surveillance system, can be partially occluded by accessories such as scarf, hat or sunglasses that make the recognition a difficult task. In this context, developing a face recognition system robust to partial occlusion means to be able to recognize people in spite of the presence of partial occlusions on the face.

There are two possible approaches to address the partial occlusion problem for face recognition. The first is to build a face recognition system which acquires good results for both the normal face and the partially occluded face. Another approach is to first pre-classify the face images into two classes with respect to face occlusion, and then customized classifiers are used for the face recognition in the two classes. The first approach is straightforward; however, such a system demands delicate tuning parameters, thus it suffers from the high complexity. On the other hand, local feature-based approaches can be integrated into the second approach, where only the non-occluded features are used for recognition. In this respect, we invest in occlusion detection methods to pre-classify the face images, in order to improve the performance of a potential face recognition system.

To the best of our knowledge, there are two main kinds of occlusion detection: the facial feature based methods [2][3] and the learning based methods [4][5]. The facial feature based methods exploit the information of facial features (such as mouth [2] or skin colour [3]) to decide whether or not a face is occluded. Whereas the learning based methods use a large number of positive (clean faces) and negative (occluded faces) samples to train a classifier, which can predict the label of an unknown face. In [4], a learning based method was proposed by combining principal component analysis (PCA) and support vector machine (SVM). In [5], the authors obtain the gradient map of the image before PCA and SVM in order to reduce the effects due to illumination variations.

The reported methods mainly tackle two different occlusions: the upper part occlusion due to sunglasses and the lower part occlusion due to scarf. To detect an occlusion caused by sunglasses is possibly less difficult since most sunglasses possess a similar appearance which is significantly different from a clean face's appearance. Furthermore, the large variation of scarf appearance with respect to structure and colour makes scarf detection a more difficult problem. Facial feature based methods [2][3] are not robust to such variations. In this paper we focus on the scarf detection problem.

Other than improving the performance of face recognition systems, scarf detection is useful in many video surveillance applications. An example could be to reinforce stadium security. It is known, in fact, that football hooligans tend to wear scarves or masks to prevent their faces being recognized before committing crimes in the stadium (see Figure 1). Hence, a scarf detection system integrated in a video surveillance system at the stadium could be useful to prevent such risks.

Figure 1 – Football hooligan wearing a scarf in a stadium (taken from internet).

 ISSN 2076-1465

Unlike regular biometric systems, surveillance systems cannot provide high quality images of the region of interest (e.g. face) in most cases. Therefore, the scarf detection should maintain a good performance for low quality images (such as noised images). In this paper we present a novel learning based method for scarf detection. Instead of using gradient map in [5], we obtain the features using Gabor wavelet before PCA and training the SVM. Experiments show that our method gives competitive accuracy in both high quality and low quality images.

2. PROPOSED METHOD

The proposed method is comprised of three parts: features extraction, dimension reduction and classification. First, features extraction is achieved by using Gabor wavelet. Then, the principal component analysis is applied to the features representation to reduce its dimension. In the end, a trained support vector machine is used to discriminate the clean and the scarf faces.

2.1 Gabor wavelet based feature extraction.

Gabor wavelets are widely used in image analysis due to their biological relevance and computational properties [6] [7]. In our method, we use Gabor wavelets to extract our features, since they can exhibit the characteristics of features in terms of spatial locality and orientation selectivity in the space and frequency domain.

2.1.1 Gabor wavelets

The first step is to build the Gabor wavelets. Gabor wavelet consists of a complex sinusoidal carrier and a Gaussian envelope. In our work, the definition of Gabor kernels is as given in [8]:

$$\mathrm{GW}_{\mu,\gamma}(\mathrm{z}) = \frac{\left\|k_{\mu,\gamma}\right\|^2}{\delta^2} e^{\left(-\|k_{\mu,\gamma}\|^2 \|\mathrm{z}\|^2/2\delta^2\right)}\left[e^{ik_{\mu,\gamma}\mathrm{z}} - e^{-\delta^2/2}\right]$$

where μ and γ are the orientation and scale of the Gabor kernels. $z = (P, Q)$ is the size of the kernel window. $\|\bullet\|$ denotes the norm operator. $k_{\mu,\gamma}$ is a wave vector defined as following:

$$k_{\mu,\gamma} = k_\gamma e^{i\Phi_\mu}$$

where

$$k_\gamma = k_{\max}/f^\gamma$$

$$\Phi_\mu = \pi\mu/8$$

k_{max} is the maximum frequency, and f is the spacing factor between kernels in the frequency domain.

In our method we set z = (20, 20), δ = 2π, k_{max}= π/2 and $f = \sqrt{2}$. To extract the features in different scales and orientations the Gabor kernels are generated in five scales γ ∈ [0, ..., 4], and eight orientations μ ∈ [0, ..., 7]. Therefore there are 40 Gabor kernels (let us indicate them with GW_i, where i ∈ [0, ..., 39]) generated for the later processing. Figure 2 shows the real part of the Gabor kernels and the corresponding magnitudes in 5 scales. In the figure, the desirable properties of spatial frequency, spatial locality and orientation selectivity is clearly shown.

2.1.2 Features Extraction

The 40 generated Gabor kernels are convoluted with the original image. Because the generated Gabor kernels are described in complex values which contain a real part GW_i^{Real} and an imaginary part GW_i^{Imag}, the two parts are convoluted with the target image separately. The process of two-dimensional convolutions with the real part and the imaginary part of Gabor kernels is described as follows:

$$\mathrm{C_i^{Real}}(x,y) = \sum_{p=0}^{P-1}\sum_{q=0}^{Q-1} \mathrm{GW_i^{Real}}(p,q)\mathrm{I}(\mathrm{x}-p, \mathrm{y}-q)$$

$$\mathrm{C_i^{Imag}}(x,y) = \sum_{p=0}^{P-1}\sum_{q=0}^{Q-1} \mathrm{GW_i^{Imag}}(p,q)\mathrm{I}(\mathrm{x}-p, \mathrm{y}-q)$$

Then the two filtered images $\mathrm{C_i^{Real}}$ and $\mathrm{C_i^{Imag}}$ are combined using a linear method:

$$\mathrm{C_i}(x,y) = \sqrt{\mathrm{C_i^{Real}}(x,y)^2 + \mathrm{C_i^{Imag}}(x,y)^2}$$

$\mathrm{C_i}$ is the filtered image which corresponds to the Gabor wavelet in a specific scale and orientation. Figure 3 shows an example image (the lower part of a clean face) and the 40 filtered images using the above mechanism. In the figure, the filtered images exhibit the properties of the original image in different scales and orientations corresponding to the Gabor wavelets shown in Figure 2.

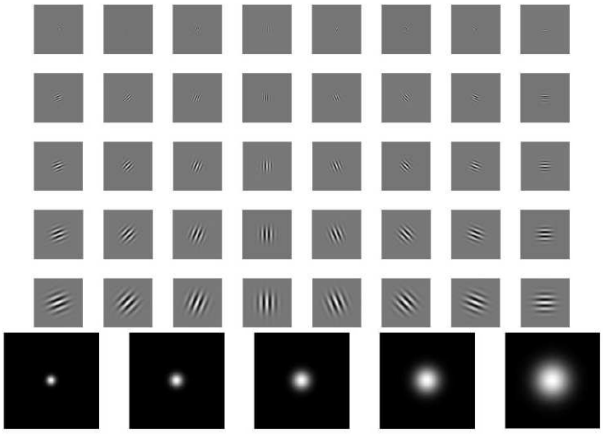

Figure 2 – Real part of the 40 Gabor wavelets and their magnitudes in five scales.

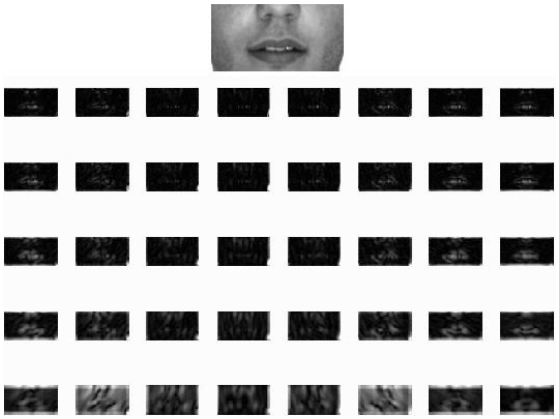

Figure 3 – An example image and the 40 filtered images by using Gabor Wavelet.

In five different scales and eight different orientations the filtered images can be described as the following set:

$$\Omega = \{C_i, i \in [0, \ldots, 39]\}$$

All 40 filtered images are used as the features for classification. An augmented feature vector is constructed by concatenating all the filtered images in Ω (row by row or column by column). Because the augmented feature vector is too large to perform the PCA, it is downsampled by a factor λ, to finally obtain the feature vector X of each image.

The selection of λ is empirical. In our experiment, λ is set to 5 according to the size of the normalized images. The feature vector X contains all the desirable features from Gabor wavelets and it is regarded as the discriminative information in the later classification task.

2.2 Principal Component Analysis

Even if the augmented feature vector is a result from the downsampling of the filtered images, the data dimension is still too high for efficient classification. In our method, the principal component analysis (PCA) is applied to the feature vectors obtained from both positive (clean faces) and negative (scarf faces) images to reduce the data dimension and to make the new representation more distinguishable.

To perform PCA we first construct a data set which consists of two types of data, in which half of the features are obtained from clean faces and another half is obtained from scarf faces. Let us define the size of the data set to be M; if X^c and X^s are the feature vectors generated from clean faces and scarf faces respectively, the data set S can be written as:

$$S = \{X_1^c, X_2^c, \dots, X_{M/2}^c, X_{M/2+1}^s, \dots, X_M^s\}$$

We obtain the mean $\bar{X}$ of the feature vectors in S by:

$$\bar{X} = \frac{1}{M}\sum_{m=1}^{M} X_m$$

The difference Φ_i between the feature vector X_i and the mean $\bar{X}$ is computed by:

$$\Phi_i = X_i - \bar{X}$$

Therefore the covariance matrix can be written as:

$$C = \frac{1}{M}\sum_{n=1}^{M} \Phi_n {\Phi_n}^T = AA^T$$

where $A = [\Phi_1, \Phi_2, \dots, \Phi_M]$. Then the eigenvectors and associated eigenvalues of the covariance matrix C can be obtained to describe the eigenspace.

All features extracted from the original image are projected onto the eigenspace.

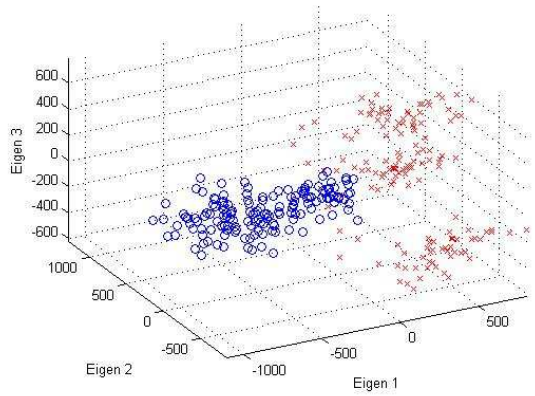

Figure 4 – Distribution of the two classes in the eigenspace.

Figure 4 shows the distribution of features in the eigenspace (projected onto first 3 eigenvectors). In the figure, the eigenspace is constructed by 150 clean faces (blue circles) and 150 scarf faces (red crosses). This observation demonstrates that the features from the two classes are roughly separated into two clusters in the eigenspace. Nevertheless, the feature projection onto all the eigenvectors is still too expensive for the computation in classification. Therefore we select the first H eigenvectors (the most discriminative ones) as the projection basis, whereby the features are obtained for classification. In our method, H is selected in order to maximize the classification accuracy during training.

2.3 Classification by Non-linear SVM

Support vector machine (SVM) is a very powerful tool for 2-class data classification in the high-dimensional space. Let's consider a training set consisting of N samples in the form $\{x_i, y_i\}_{i=1}^N$, in which x_i is the feature vector of a sample and $y_i \in \{-1,1\}$ is the label which indicates which class x_i belongs to. SVM finds the maximum-margin hyperplane to separate the data by:

$$f(x_i) = sign\left(\sum_{j=1}^{N} \alpha_i y_i K(x_i, x_j) + b\right)$$

where $\{x_j, j \in [1, N]\}$ are the support vectors. Non-linear SVM applies the "kernel trick" to fit the maximum-margin hyperplane in a transformed feature space. Here we use the Radial Basis Function (RBF) kernel. The RBF kernel can be written as:

$$K(x_i, x_j) = e^{-\gamma\|x_i - x_j\|^2}, \gamma > 0$$

In our experiments, the tool LIBSVM [9] is used to perform the non-linear SVM.

3. EXPERIMENTS AND RESULTS

In our experiments, the proposed method is compared with the two other learning methods. Because our method uses Augmented feature vector, PCA and SVM, we give it a short name as APS. Similarly, the methods presented in [4] and [5] are named as PS (PCA and SVM) and GPS (Gradient map, PCA and SVM) in the same manner.

3.1 Databases

The first database we used is the AR face database (ARFD) [10]. It is widely used in the experiments for face detection and face recognition related to the occlusion problem because it contains a large number of occluded faces. Figure 5 shows four examples of face images from ARFD.

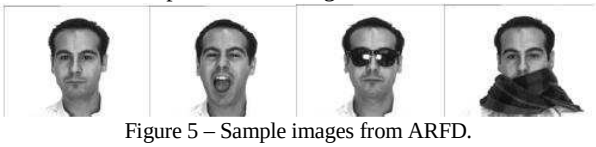

Figure 5 – Sample images from ARFD.

In ARFD, 300 clean faces and 300 scarf faces are selected. The clean faces are comprised of the faces of 50 males and 50 females with 3 different facial expressions (natural expression, smile and anger); where the scarf faces are comprised of the faces of 50 males and 50 females with scarves

under 3 different illuminations (left side light, right side light and all sides light).

In ARFD, most of the scarves are dark, and there is a lack of the variation of structures in the scarf appearance. In real world, the scarves can have various colours (even lighter than the skin colour) and various structures on the appearance (e.g. grids on the scarf). To test the robustness of the methods for various scarf appearances, we built the second face database – the EURECOM face database (ERFD).

ERFD consists of 72 face images taken from 6 lab members in 2 sessions. In each session, there are 3 clean faces with different facial expressions (natural expression, smile and talking) and 3 scarf faces with different scarf appearances for each individual. We select the clean faces with different facial expressions in order to increase the variation of each individual, so that images are more suitable to simulate the real conditions from video surveillance system. In total there are 36 clean faces and 36 scarf faces. Figure 6 shows the scarf faces extracted from ERFD. It exhibits the large variation of scarf appearance.

Figure 6 – Examples of scarf face from ERFD.

For both databases, the face region was extracted from the original images. Then the lower half of the face region was segmented for scarf detection. All the extracted lower part of faces was normalized into a fixed size of 90*50 pixels.

3.2 Eigenvectors selection

PCA is the essential step in PS, GPS and APS to obtain the final features. We select the first H eigenvectors as the projection basis. The choice of H for each method is empirical. In our experiments we select H to maximize the classification accuracy of each method during training as shown in the following results.

3.3 Experiments on ARFD

We define the positive norm as scarf face (i.e. a clean face classified as a scarf face is a false acceptance whereas a scarf face classified as a clean face is a false rejection). The results of PS, GPS and APS on ARFD are shown in Table1.

Method	FAR	Detection Rate	H
PS	2.67%	99.33%	38
GPS	2%	98.67%	29
APS	1.33%	99.33%	51

Table 1 – Results of PS, GPS and APS on ARFD.

From the results in the table we can observe that the proposed method (APS) gives very good detection rate with the lowest false acceptance rate. The classification accuracy of the three methods on various numbers of eigens (1-100) is shown in Figure 7. It shows that the proposed method (APS) outperforms the others when using a higher number of eigens.

3.4 Experiments on ERFD

The results of PS, GPS and APS on ERFD are shown in Table 2. Notice that the ERFD is relatively small (only 18 clean faces and 18 scarf faces are used for training). Therefore the variation of statistical results in ERFD is greater than in ARFD.

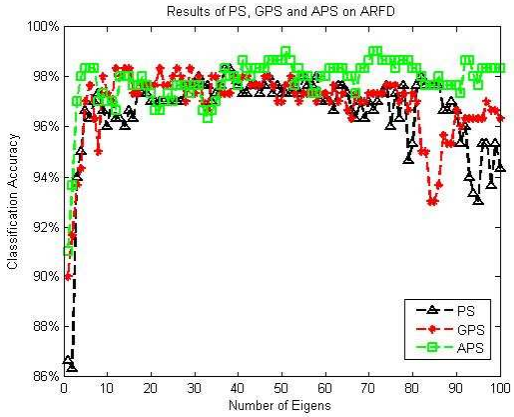

Figure 7 – Classification accuracy of PS, GPS and APS on ARFD.

Method	FAR	Detection Rate	H
PS	11.11%	77.78%	7
GPS	5.55%	100%	22
APS	0%	100%	7

Table 2 – Results of PS, GPS and APS on ERFD.

From the results in the table we can see that both GPS and APS give the perfect detection rate. But the false acceptance rate in GPS is higher than in APS. Figure 8 shows the classification accuracy of the three methods on various numbers of eigens (1-36). In the figure, both GPS and APS provide good performances. In contrast, the accuracy of PS is significantly lower than GPS and APS. Therefore, directly using the intensity image for classification (PS) is not robust to the variation of scarf appearance.

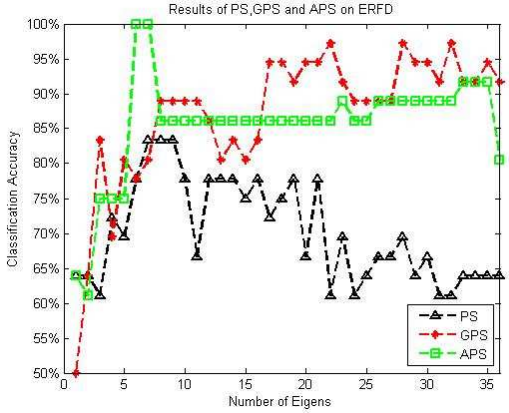

Figure 8 – Classification accuracy of PS, GPS and APS on ERFD.

3.5 Experiments on Gaussian noised images from ERFD

In video surveillance systems, the image quality may not be optimal as the images obtained from laboratories. Hence the scarf detection is required to be robust to noise. We tested our method (APS), PS and GPS on a Gaussian noised version of images from ERFD. In the experiments, we first add a Gaussian noise to the original images (here the Gaussian noise is zero mean with the variance at 0.005). Then we applied the different methods to the images after noising. Figure 9 shows the classification accuracy of PS, GPS and APS on the Gaus-

sian noised images. In the figure, APS maintains good classification accuracy whereas the accuracy of GPS significantly decreases comparing to the results in Figure 8.

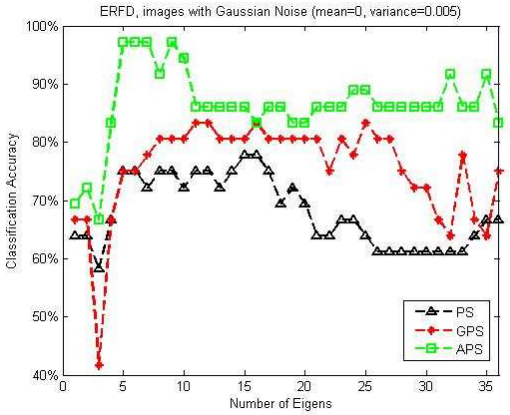

Figure 9– Classification accuracy of PS, GPS and APS on Gaussian noised (mean=0, variance=0.005) images from ERFD.

In order to prove that our method is more robust than PS and GPS with different strengths of noise, we vary the strength of Gaussian noise in the experiment. The mean of Gaussian noise is set to be zero and the variance of Gaussian noise is varying from 0.005 to 0.05 by step of 0.005. Figure 10 shows the effect of Gaussian noise with different strengths.

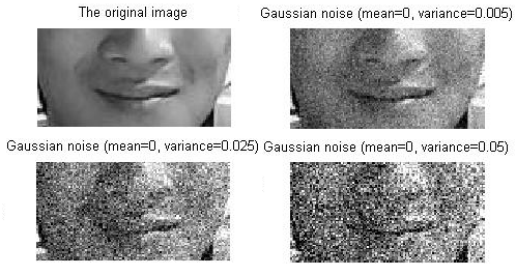

Figure 10 – Examples of Gaussian noised image with various strengths.

Figure 11 shows the result of this experiment. Here the number of eigens used in each method is fixed ($H = 20$). In the figure the APS maintains good classification accuracy for all levels of strengths. In contrast, the PS and GPS performances dramatically decrease when the noise strength increases.

4. CONCLUSIONS

In this paper we have presented a new learning based method for efficient scarf detection. The proposed method (APS) is compared with state of art methods (PS and GPS) through various experiments. Both APS and GPS are efficient solutions in restricted environments (e.g. biometric systems). But only APS qualifies the desirable performance in unconstrained environments (e.g. video surveillance systems).

The next step is to study how to combine our method with face recognition in order to properly address the face occlusion problem. Another research direction could be to extend the current method from still images to video, in which the human behaviour of putting on/off a scarf should also be analyzed.

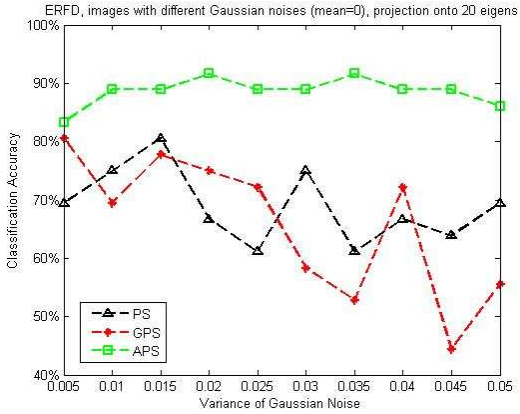

Figure 11–Classification accuracy of PS, GPS and APS on Gaussian noised images, mean=0, variance=0.005 to 0.05, H=20.

ACKNOWLEDGMENT

The research described in this paper has been partially funded by the FR OSEO BIORAFALE project.

REFERENCES

[1] W. Zhao, R. Chellappa, P. J. Phillips, A. Rosenfeld, "Face Recognition: A Literature Survey," *Computing Surveys*, vol. 35, part 4, pp. 399-459, 2003.

[2] C.-Y. Wen, S.-H. Chiu, Y.-R. Tseng and C.-P. Lu, "The Mask Detection Technology for Occluded Face Analysis in the Surveillance System, " *Journal of forensic sciences*, vol. 50, no. 3, pp. 593-601, 2005.

[3] D.-T. Lin, M.-J. Liu, "Face Occlusion Detection for Automated Teller Machine Surveillance," *Lecture Notes in Computer Science*, vol. 4319, pp. 641-651, Sept. 2006.

[4] S. M. Yoon, S. C. Kee, "Detection of Partially Occluded Face using Support Vector Machines," in *Proc. IAPR Work Mach Vis Appl 2002*, Japan, 2002, pp. 546-549.

[5] J. Kim, Y. Sung, S.M. Yoon, and B.G. Park, "A new video surveillance system employing occluded face detection," *Lecture Notes in Computer Science*, page 68, 2005.

[6] C. Liu and H. Wechsler, "Gabor feature based classification using the enhanced Fisher linear discriminant model for face recognition," *IEEE Trans. on Image Processing*, vol. 11, part 4, pp.467–476, Apr. 2002.

[7] Du, Shan and Ward, Rabab Kreidieh, "Improved face representation by nonuniform multilevel selection of Gabor convolution features," *Trans. Sys. Man Cyber. Part B*, vol. 39, no. 6, pp. 1408--1419, 2009.

[8] M. Lades, J.C. Vorbrueggen, J.M. Buhmann, J. Lange, C. Malsburg, R.P. Wuertz, and W. Konen. "Distortion invariant object recognition in the dynamic link architecture," *IEEE Transactions on computers*, vol. 42, part 3, pp.300–311, 1993.

[9] Chih-Chung Chang and Chih-Jen Lin, "LIBSVM : a library for support vector machines", 2001. Software available at http://www.csie.ntu.edu.tw/~cjlin/libsvm.

[10] A.M. Martinez and R. Benavente, " The AR face database," Technical report, CVC Technical report, no. 24, 1998