

Precise eye localization through a general-to-specific model definition

Paola Campadelli*, Raffaella Lanzarotti, Giuseppe Lipori
Dipartimento di Scienze dell'Informazione
Università degli Studi di Milano
Via Comelico, 39/41 20135 Milano, Italy
{campadelli, lanzarotti, lipori}@dsi.unimi.it

Abstract

We present a method for precise eye localization that uses two Support Vector Machines trained on properly selected Haar wavelet coefficients. The evaluation of our technique on many standard databases exhibits very good performance. Furthermore, we study the strong correlation between the eye localization error and the face recognition rate.

Keywords - Eye localization, face recognition, Haar wavelets, Support Vector Machines (SVM).

1 Introduction

In the last two decades researchers devoted a great effort to the development of Face Recognition (FR) systems; nowadays the interest in the topic is still high, but the focus is on a bigger challenge: the objective is to find solutions adequate to real-world applications, that is completely automatic and robust in unconstrained contexts [28].

In this framework one of the critical unsolved problem regards the precise and automatic localization of facial features (eyes, nose and mouth): the FR methods require either to warp the input images, or at least to normalize them. This makes the facial feature localization particularly crucial, above all considering that even small localization errors make the recognition process fail [19, 20].

Many recent papers (e.g. [5, 6, 18, 14, 17, 22, 26, 27]) tackle some of the critical aspects arising in uncontrolled environments (variations of the illumination, changes of the face pose and expression, eventual partial occlusions), however they do not deal with the feature localization problem. Either implicitly or explicitly, they refer for their initialization to manual annotations of the facial features, so compromising the automation of the FR systems and thus limiting the applicability in real-world contexts.

Indeed, some researchers have faced the problem: the first were Jesorsky et al. [16], Smeraldi et al. [23] and Martínez [20], and recently there has been a new impulse in this very relevant research field [7, 13, 15, 19, 25], while leaving room for further improvements.

*Work partially supported by the PASCAL Network of Excellence under EC Grant no. 506778 and by the FIRST grant of the University of Milan. This publication only reflects the authors view.

In this paper we present in details our eye localization method, showing its performance on standard databases (XM2VTS, BANCA, FRGC, FERET, and BioID) and comparing it with the algorithms which report results using the distance measure introduced in [16]. Like other researchers, we focus on the eye localization: once the eyes have been precisely localized the position of the other features can be easily determined.

The method assumes to receive as input a rough estimation of the face position and scale, obtained by any face detector. We model the eye at two different levels of detail. The first module (*eye detector*) is used to both validate the regions found by the face detector and to roughly localize the eyes. The second module (*eye localizer*) works in cascade and determines the precise position of the eyes center. Both modules require the application of the Haar wavelet transform, and a procedure to select proper sets of coefficients which characterize the patterns robustly. The two sets of Haar wavelet coefficients are used to train two Support Vector Machines (SVMs).

The rest of the paper is organized as follows: section 2 describes the pattern definition; the proposed methodology is sketched in section 3, while in section 4 extensive results are reported and compared with the best algorithms presented in the literature; section 5 draws the conclusions of our work.

2 Pattern definition and feature selection

The difficulty intrinsic to the task of eye localization requires an accurate choice of a suitable representation of the eye pattern. In [3, 4] we have observed that the wavelet representation is more favorable than the direct representation as it leads to a smaller generalization error. Haar-like wavelets permit to describe the pattern in terms of luminance changes at different frequencies, at different positions and along different orientations.

Before the wavelet decomposition, each image undergoes an illumination normalization process and is then reduced to 16×16 pixels¹. The decomposition is realized via an *overcomplete* bi-dimensional FWT (Fast Wavelet Transform) [3]. A wavelet coefficient $d_{j,k1,k2}^o$ is identified by four parameters: j is called the detail level and relates to the size of the window over which the coefficient is calculated (hence it regulates the frequency); $(k1, k2)$ is called the shift and relates to the position of the coefficient within the image; $o \in \{horizontal, vertical, diagonal\}$ determines the orientation of the edge that is tested for presence. The set B_j of all $d_{j,k1,k2}^o$ of a certain level j is called the band of level j .

The scaling coefficients $c_{j,k1,k2}$ describe the mean illumination of the image; we discard them producing a sort of illumination normalization of the pattern examples. A further manual selection can be done by observing that the band corresponding to the the highest frequency is crucial for the precision of the eye localizer, while it can be safely discarded in the case of the eye detector since it must be general and robust even if at the expenses of a coarser localization. By doing so we manage both to specialize the two classifiers by specifying a different maximum detail level, and to exclude *a priori* a big number of coefficients when they are not needed ($|B_0| = 675$, after their exclusion we are left with $\sum_{j=1}^3 |B_j| = 177$ coefficients).

In order to carry out the real selection process, we need a way to assess the relative importance of the coefficients left. We do this following the idea proposed in [21] by

¹Such a dimension represents a trade off between the necessity to maintain low the computational cost and to have sufficient details to learn the pattern shape.

applying a normalization step: we take a set L of eye pattern images ($|L| = 2152$ in our case) and we decompose each $l \in L$ in its wavelet coefficients $d_{j,k_1,k_2}^o(l)$. For each coefficient d_{j,k_1,k_2}^o we calculate its mean value in the sample of patterns:

$$\overline{d}_{j,k_1,k_2}^o = \frac{\sum_{l=1}^{|L|} |d_{j,k_1,k_2}^o(l)|}{|L|}$$

and we normalize it with respect to the average mean of its band; then

$$\tilde{d}_{j,k_1,k_2}^o = \frac{\overline{d}_{j,k_1,k_2}^o}{m_j}, \quad \text{where } m_j = \frac{\sum_{k_1} \sum_{k_2} \sum_o \overline{d}_{j,k_1,k_2}^o}{|B_j|},$$

represent the normalized coefficients that can be now ordered to assess their relative importance. Notice that the normalization is done here within the entire band and not within each sub-band (a subset of a band whose coefficients have all the same orientation).

Since the expected value of the sum of all \tilde{d}_{j,k_1,k_2}^o in the same band is approximately equal to their cardinality²

$$E \left[\sum_{k_1} \sum_{k_2} \sum_o \tilde{d}_{j,k_1,k_2}^o \right] = \sum_{k_1} \sum_{k_2} \sum_o E \left[\frac{\overline{d}_{j,k_1,k_2}^o}{m_j} \right] \approx \frac{\sum_{k_1} \sum_{k_2} \sum_o E[\overline{d}_{j,k_1,k_2}^o] \cdot |B_j|}{E[\sum_{k_1} \sum_{k_2} \sum_o \overline{d}_{j,k_1,k_2}^o]} = |B_j|$$

the normalized coefficients $\tilde{d}_{j,k_1,k_2}^o > 0$ can be interpreted as follows:

$$\tilde{d}_{j,k_1,k_2}^o \begin{cases} \sim 1 & \Rightarrow \text{no regularity} \\ \ll 1 & \Rightarrow \text{systematic uniformity} \quad (C^-) \\ \gg 1 & \Rightarrow \text{systematic variation} \quad (C^+) \end{cases}$$

Hence the normalization allows us to distinguish two sub-categories of coefficients that can be ordered separately: C^+ , the coefficients that are systematically greater than 1, and C^- , those which are systematically smaller than 1. Both of them retain precious information: the first class gathers the coefficients that capture the edge structure of the pattern, while the second class contains the coefficients that indicate a systematic absence of edges (in a certain position, at a certain frequency and along a certain orientation).

Once ordered the normalized coefficients, we define an error function to drive the selection process. We can measure the expressiveness of the coefficients by measuring how well they reconstruct the pattern that they represent (i.e. the mean eye). We wish to find

$$w = \arg \min_{\substack{w = w^+ \cup w^-, \\ w^+ \subseteq C^+, w^- \subseteq C^-}} \|E - E_w\|^2 + \alpha \cdot \|E_w - U\|^2$$

where E is the mean eye pattern, U is a uniform pattern (with all pixels set to the mean luminance of E) and E_w is the reconstruction obtained by retaining the set w of the wavelet coefficients $w^+ \subseteq C^+$ and $w^- \subseteq C^-$. The first term of the objective function represents the error made by the reconstruction, while the second term intends to bound the amount of detail we are adding to the pattern representation (the value α is a trade-off to balance

²The equality does not hold strictly because the random variables $\overline{d}_{j,k_1,k_2}^o$ and m_j are not independent. However they can be assumed to be so, as their correlation is very low (m_j depends on the sum of many $\overline{d}_{j,k_1,k_2}^o$).

between these two opposite goals). We can exploit the ordering of the coefficients and thus we avoid to optimize over all the possible subsets of $C^+ \cup C^-$: w is incremented by iteratively adding new coefficients according to their ordering.

We experimentally observed that the trend of the objective function is not very sensitive to small variations of α in the interval $[0.5, 1]$; we set it to 0.8. As it can be expected, the norm of the reconstruction maximally varies as we increase the number of w^+ retained, while it is almost insensitive to the number of selected w^- . We select the $w = w^+ \cup w^-$ such that it corresponds to a local minimum of the objective function, and the ratio $|w^+|/|C^+|$ roughly equals $|w^-|/|C^-|$.

Figure 1 shows the coefficients selected for the pattern characterization of each classifier. For the eye detector we retain less than 100 wavelet coefficients that well characterize the general eye shape. The representation associated with the eye localizer keeps about 300 coefficients, therefore the application of the second classifier is more costly than the first one. The differentiation of the two SVMs is also achieved by choosing a suitable

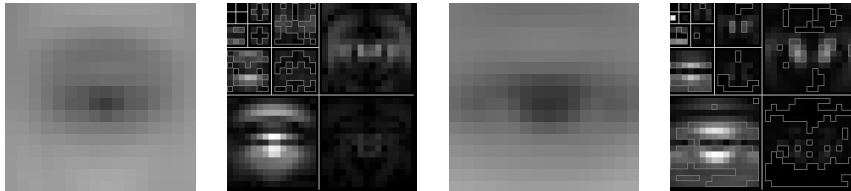


Figure 1: From left to right: the mean eye pattern, its wavelet decomposition and the selected features (red contour) for the SVM-1 and the SVM-2 respectively. High intensities correspond to strong edges, low intensities indicate uniform regions.

sample of training examples for each of them: the SVM-1 (eye detector) must distinguish the global eye shape from that of other patterns, especially those found inside the face. The positive class is built to contain eye images cropped to a size equal to the inter-ocular distance. The negative class is populated by the other facial features (nose, mouth, chin, cheeks, forehead, etc.) and by some examples extracted from the background of the images (respectively 3 and 2 for every positive). The SVM-2 (eye localizer) is presented with positive examples that correspond to a smaller receptive field (half of the eye pattern previously defined) and with negative examples that are generated by small, random displacements of the sub-images used for the extraction of the positive examples (10 negative examples for each positive).

The training and test sets for both SVMs are derived from the same set of images: 600 from the FERET database [11] (controlled images of frontal faces), 400 from the BANCA database [8] (to model the different illumination conditions and the closed eyes), and 600 from a custom database containing many heterogenous and uncontrolled pictures of various people (useful to model pose variations and non-neutral face expressions).

3 Localization technique

Although we assume to receive as input the information regarding the position and the scale of the face in the image, we know that any automatic face detector is subject to a certain error distribution on the size of its outputs, besides the presence of possible false

positives. We account for the first uncertainty by considering a range of three scales at which to search for eyes, and we use the eye detector as a “validator” of the face found by the detector to discard possible false positives.

Given a region output by the face detector, the evaluation of a candidate point P within that region comes to evaluating three examples centered in it: the one at the inferred scale (example \mathbf{x}_P), plus two examples extracted according to a small underestimation and a small overestimation of that scale (examples \mathbf{x}_P^- and \mathbf{x}_P^+). The size of \mathbf{x}_P^- and \mathbf{x}_P^+ are chosen to account for an estimation of the face size that is between half and twice the true size.

If we call $SVM(\mathbf{x}) = 0$ the equation of the hyperplane that separates the two classes of positive and negative examples, then we can treat $SVM(\mathbf{x})$ as a “measure” of the confidence with which the SVM classifies the example \mathbf{x} . Thus we define the function

$$\rho(P) = SVM_1(\mathbf{x}_P) + SVM_1(\mathbf{x}_P^-) + SVM_1(\mathbf{x}_P^+)$$

as the strength of the candidate point P .

We proceed by evaluating $\rho(P)$ over a small subset of points in the face region: first we identify the points that lie on the edges, then we subsample them with a step that depends on the scale of the face region; we consider as point candidates the ones for which $\rho(P)$ is greater than 0, and we group them according to their proximity in the image; each group of point candidates is then represented by its centroid (the eye candidate) obtained weighting each point P with its $\rho(P)$. This last step strengthens the eye detection, making it more stable.

At last, we refine the results of the detection by applying the SVM-2 within a small neighborhood of the found positions. The scale considered by the eye localizer should approximate half of the inter-ocular distance and is obtained as follows:

$$\frac{1}{2} \times \frac{\sum_{\mathbf{x} \in \{\mathbf{x}_P, \mathbf{x}_P^+, \mathbf{x}_P^-\}} [\Theta(SVM_1(\mathbf{x})) \times scale(\mathbf{x})]}{3} \quad \text{where} \quad \Theta(z) = \begin{cases} z & \text{if } z > 0 \\ 0 & \text{if } z \leq 0 \end{cases}$$

This quantity exploits the responses of the SVM-1 as it weights the three different scales according to their margins (if positive) in order to get an estimate that is more robust to possible errors of underestimation or overestimation of the face size. The search for the candidates that give the highest response according to the SVM-2 proceeds analogously as in the eye detection module.

4 Experimental results

In order to assess the precision of the eye localization we adopt the relative error measure proposed by Jesorsky [16]:

$$d_{eye} = \frac{\max(\|C_l - \tilde{C}_l\|, \|C_r - \tilde{C}_r\|)}{\|C_l - C_r\|}$$

where the values $\tilde{C}_{r/l}$ stand for the eye positions output by the localization module, while the values $C_{r/l}$ are the ground truth of the right and left eye centers respectively. This measure, which normalizes the localization error over the inter-ocular distance, is scale

independent and therefore it permits to compare datasets characterized by different resolutions of the face region. We refer to it for the quantitative evaluation of both methods of eye detection and eye localization. Regarding the former, we observe that there is a general agreement [16, 19] that $d_{eye} \leq 0.25$ is a good criterion to identify correct eye detection. Regarding the latter, following the idea presented in [19], we studied the relationship between d_{eye} and the face recognition rate of some baseline methods available in the CSU package [2] and our algorithm presented in [1].

To mimic the behavior of eye localization techniques that achieve different levels of precision, we carried out four recognition experiments by increasingly perturbing the ground truth quality; both C_r and C_l have been randomly displaced inside circles of radii 0.05, 0.1 and 0.15 with uniform distribution. In Figure 2 we report the results of this study on the XM2VTS database [12], which contains 1180 pictures grouped into 4 sessions of 295 subjects. The experiment is defined as follows: session 1 is used for the gallery, session 2 for the probe, sessions 3 and 4 constitute the training set. All three sets (gallery, probe and training) have been perturbed as it would happen in a completely automatic system. We note here that our algorithm does not need a training phase.

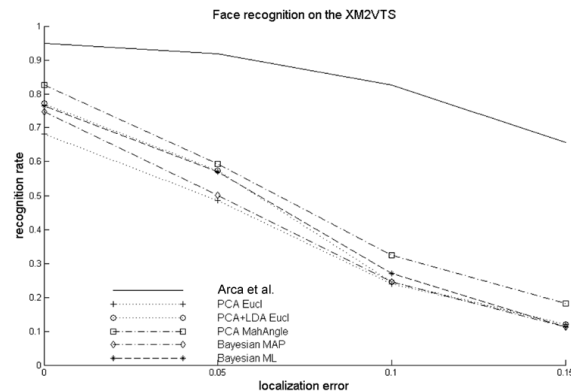


Figure 2: Face recognition vs. eye localization precision. All methods based on PCA maintain 107 eigenvectors to build the feature subspace.

The graph of Figure 2 clearly shows that the precision of eye localization is extremely critical for the methods which require the accurate alignment of faces (like the subspace projection techniques). In particular the baseline methods suffer from a drop of the face recognition rate by about the 20% even for small eye localization errors ($d_{eye} = 0.05$). In such context 0.05 should be considered as the maximum acceptable error, as noted by Hamouz et al. [15]. Recently a similar result appeared in the literature [25], where an even more drastic drop of performance is presented (50%, at $d_{eye} = 0.05$), obtained on the FRGC database and running a different implementation of the PCA technique.

In Figure 3 we present the results of the application of our localization technique to many different datasets in terms of receiver operator curves (ROC). Besides the already mentioned XM2VTS, the experiments have been carried on: 416 pictures from the BANCA database of English people (208 from the *adverse* section, 208 from the *controlled* one); 375 pictures taken from the FERET dataset (200 ba, 175 fb); 1521 pictures from the BioID dataset [9]; 862 images from the FRGC database [10] (half taken in con-

trolled conditions, half in uncontrolled conditions). All these subsets contain images of single faces with vertical, frontal pose, eyes closed or open, presence or absence of spectacles; none of these images has been used for the training of the SVM classifiers. The ROC curves tell the rate of the automatic eye localization for each level of precision d_{eye} as measured with respect to the ground truth of the eye centers.

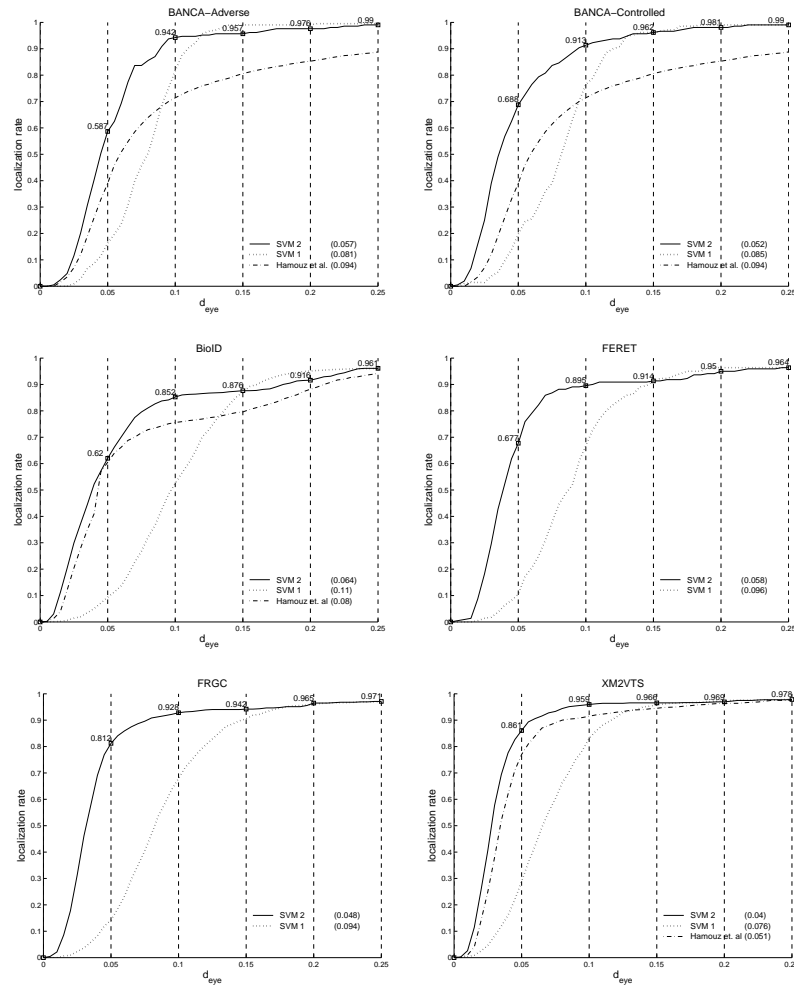


Figure 3: The ROC curves of eye detection and localization over different databases.

The graphs compare the performance of our eye detector (SVM-1) against our eye localizer (SVM-2) and, where available, we report the performance of the method [15] denoted as ‘1 face on the output’. Adopting $d_{eye} \leq 0.25$ as the criterion for correct eye detection, we observe that the SVM-1 alone permits to achieve rates of 99.0%, 96.1%, 96.4%, 97.1% and 97.8% over the datasets BANCA, BioID, FERET, FRGC and XM2VTS respectively. As expected, the addition of the second classifier greatly improves the precision of the detection and the curves are systematically above the rates declared by Hamou et al. The performance of the SVM-2 evaluated at $d_{eye} \leq 0.05$ achieves rates of 58.7%,

68.8%, 62.0%, 67.7%, 81.2% and 86.1% respectively. The numbers reported in parenthesis on the graphs represent the surface area enclosed between each ROC and the perfect localization (rate identically equal to 1), therefore they give a global evaluation of the localization performance of the method over that particular dataset (the smaller, the better).

To our knowledge [16], [19] and [24] are the only other works that use the error measure d_{eye} in order to assess the quality of eye localization. The former exhibits a localization performance that is lower than that reported by Hamouz and others. The second presents a ROC curve that looks similar to the performance of our SVM-1 but it is obtained referring to a mix of databases with no intersection with the ones we considered, making impossible a direct comparison. Finally, the third paper reports results on the BioID and on the JAFFE databases, focusing only on $d_{eye} \leq 0.1$ and $d_{eye} \leq 0.25$. The performance reported on the BioID are better than ours (91.8% and 98.1% respectively), but we do not know how the method behaves at $d_{eye} \leq 0.05$.

Other works face the same problem, while adopting a different metric: Wang and others [25] adopt a normalized mean error (not the maximum, which is a worst case analysis) and give an error of 2.67% on the entire FRGC. By adopting this measure on our experiments, we observe an error of 2.65% and 3.88% on the controlled and uncontrolled images of our experiments (respectively). If we consider that the entire FRGC contains twice as many controlled images as uncontrolled ones, we see that these results are very similar. Also the work [7] adopts a relative error metrics m_e but it averages the localization error of seventeen facial features. This is a much stronger objective than localizing only the eyes, however the global rate reported over the BioID is significantly lower than our. Since it is impossible to extrapolate the precision of the sole eye localization, a direct comparison cannot be made.

We strongly support the use of the d_{eye} error measure in order to make real comparisons of the eye localization performance of different methods.

5 Summary and conclusions

In this paper we presented a new method to precisely localize the eyes. The method consists of subsequent refinement steps; assuming to have roughly localized the faces in the image, we apply an eye detector with the aim of discriminating between real face regions and possible false positives. If a more precise eye localization is required, we apply an eye localizer that works on a small neighborhood of the positions output by the detector. The experiments have been carried out on a wide set of images extracted from databases with different characteristics. The quality of the detection is near optimal and the localization results improve those reported by the algorithms presented in the literature. However the localization rates might not be sufficient for algorithms which require almost pixelwise alignment. In this respect, we think that it would be very useful if the description of face recognition methods included the required alignment precision.

Regarding the computational time, the eye detection module is much faster than the eye localization: the mean execution time on faces with an inter-ocular distance of 70 pixels is about 4 sec for eye detection, and three times as much for localization (java interpreted code running on a Pentium 4 with 3.2GHz). The speed was not the focus of our research but we plan to optimize the implementation by substituting the direct FWT calculation of the Haar coefficients with the adoption of the integral image calculation.

References

- [1] S. Arca, P. Campadelli, and R. Lanzarotti. A face recognition system based on automatically determined facial fiducial points. *Pattern Recognition*, 39(3):432–443, 2006.
- [2] J.R. Beveridge, D. Bolme, B.A. Draper, and M. Teixeira. The CSU face identification evaluation system. its purpose, features, and structure. *Machine vision and applications*, 16:128–138, 2005.
- [3] P. Campadelli, R. Lanzarotti, and G. Lipori. Face localization in color images with complex background. *Proceedings of the IEEE International Workshop on Computer Architecture for Machine Perception (CAMP 2005), Palermo, Italy.*, pages 243–248, 2005.
- [4] P. Campadelli, R. Lanzarotti, and G. Lipori. Eye localization and face recognition. *RAIRO, to be published*, 2006.
- [5] H. Cevikalp, M. Neamtu, M. Wilkes, and A. Barkaba. Discriminative common vectors for face recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 27(1):4–13, 2005.
- [6] K.I. Chang, K.W. Bowyer, and P.J. Flynn. An evaluation of multimodal 2d+3d face biometrics. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 27(4):619–624, 2005.
- [7] D. Cristinacce and T.F. Cootes. A comparison of shape constrained facial feature detectors. *Proceedings International conference on Automatic Face and Gesture Recognition*, 2004.
- [8] The BANCA Database.
Web address: <http://www.ee.surrey.ac.uk/Research/VSSP/banca/>.
- [9] The BioID Database.
Web address: <http://www.humanscan.de/support/downloads/facedb.php>.
- [10] The Face Recognition Grand Challenge Database.
Web address: <http://www.frvt.org/FRGC/>.
- [11] The FERET Database.
Web address: <http://www.itl.nist.gov/iad/humanid/feret/>.
- [12] The XM2VTS Database.
Web address: <http://www.ee.surrey.ac.uk/Research/VSSP/xm2vtsdb/>.
- [13] I. Fasel, B. Fortenberry, and J. Movellan. A generative framework for real time object detection and classification. *Computer Vision and Image Understanding*, 98:182–210, 2005.
- [14] G. Givens, J.R. Beveridge, B.A. Draper, P. Grother, and P.J. Phillips. How features of the human face affect recognition: a statistical comparison of three face recognition algorithms. *Proceedings of IEEE on Computer Vision and Pattern Recognition*, 2004.

- [15] M. Hamouz, J. Kittler, J.K. Kamarainen, P. Paalanen, H. Kälviäinen, and J. Matas. Feature-based affine invariant localization of faces. *IEEE transactions on pattern analysis and machine intelligence*, 27(9):1490–1495, 2005.
- [16] O. Jesorsky, K.J. Kirchberg, and R.W. Frischholz. Robust face detection using the hausdorff distance. *Lecture Notes in Computer Science*, 2091:212 – 227, 2001.
- [17] J. Kim, J. Choi, J. Yi, and M. Turk. Effective representation using ica for face recognition robust to local distortion and partial occlusion. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 27(11):1977–1981, 2005.
- [18] C. Liu. Gabor-based kernel PCA with fractional power polynomial models for face recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 26(5):572–581, 2004.
- [19] Y. Ma, X. Ding, Z. Wang, and N. Wang. Robust precise eye location under probabilistic framework. *Proceedings of IEEE International Conference on Automatic Face and Gesture Recognition (FGR 04)*, 2004.
- [20] A.M. Martinez. Recognizing imprecisely localized, partially occluded, and expression variant faces from a single sample per class. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(6):748–763, 2002.
- [21] C.P. Papageorgiou and T. Poggio. Trainable pedestrian detection. In *Proceedings of International Conference on Image Processing*, pages 35–39, 1999.
- [22] S. Romdhani and T. Vetter. Estimating 3d shape and texture using pixel intensity, edges, specular highlights, texture constraints and a prior. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR 05)*, 2005.
- [23] F. Smeraldi and J. Bigun. Retinal vision applied to facial features detection and face authentication. *Pattern recognition letters*, 23:463–475, 2002.
- [24] X. Tang, Z. Ou, T. Su, H. Sun, , and P. Zhao. Robust precise eye location by adaboost and svm techniques. *Proceedings of the International Symposium on Neural Networks*, pages 93–98, 2005.
- [25] P. Wang, M. Green, Q. Ji, and J. Wayman. Automatic eye detection and its validation. *Proceedings of the Workshop FRGC in the IEEE conference on Computer Vision and Pattern Recognition*, 2005.
- [26] X. Wang and X. Tang. A unified framework for subspace face recognition. *IEEE Transactions Pattern Analysis and Machine Intelligence*, 26(9):1222–1228, 2004.
- [27] B. Zhang, X. Chen, S. Shan, and W. Gao. Nonlinear face recognition based on maximum average margin criterion. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR 05)*, 2005.
- [28] W. Zhao, R. Chellappa, P.J. Phillips, and A. Rosenfeld. Face recognition: A literature survey. *ACM, Computing Surveys*, 35(4):399–458, 2003.