Eyes Closeness Detection
Using Appearance Based Methods

Liu Xue, Tan Xiaoyang, and Chen Songcan

College of Computer Science & Technology, Nanjing University of Aeronautics &
Astronautics, P.R.China
{liuxue,x.tan,s.chen}@nuaa.edu.cn

Abstract. Human eye closeness detection has gained wide applications
in human computer interface designation, facial expression recognition,
driver fatigue detection, and so on. In this work, we present an exten-
sive comparison on several state of art appearance-based eye closeness
detection methods, with emphasize on the role played by each crucial
component, including geometric normalization, feature extraction, and
classification. Three conclusions are highlighted through our experimen-
tal results: 1) fusing multiple cues significantly improves the performance
of the detection system; 2) the AdaBoost classifier with difference of in-
tensity of pixels is a good candidate scheme in practice due to its high
efficiency and good performance; 3) eye alignment is important and in-
fluences the detection accuracy greatly. These provide useful lessons for
the future investigations on this interesting topic.

Keywords: Eye closeness detection, Eye state measurement

1 Introduction

As one of the most salient facial features, eyes, which reflect the individual's
affective states and focus attention, have become one of the most important
information sources for face analysis. Efficiently and accurately understanding
the states of eyes in a given face image is therefore essential to a wide range
of face-related research efforts, including human computer interface designation,
facial expression analysis, driver fatigue detection [1][2], and so on.

However, this task is challenging due to the fact that the appearance of eye
regions can be easily influenced by various variations such as lighting, expression,
pose, and human identity. To meet these challenges, numerous eye closeness de-
tection methods have been proposed during past few decades [3][4][5][6][7][8][9].
The ideas of these methods can be roughly categorized into two types, i.e., detect-
ing the closeness of eyes directly through various pattern recognition methods
or doing this indirectly by checking whether the eyes are actually open. Since
an open eye in general exhibits more appearance evidence (e.g., visible iris and
elliptical shape of eyelids), most methods (e.g., [6][7][8]) understand the state of
the eyes using this later philosophy, i.e., if the local evidence collected from the
image supports the conclusion that the eyes are open, then they must be not

closed. However, methods in this line may have their own problems. For example, for some subjects, their irises may be largely occluded by their eyelids. As a result, it would be very hard to reliably detect irises under this situation [5].

In this paper, we adopt the first philosophy, i.e., detecting whether the eyes are closed directly using their appearance evidence. The major advantage of this lies in its efficiency and robustness [10][11]. Current development in computer vision has allowed for robust middle-level feature description for eye patches despite of various changes in appearance, and the remaining variations can be addressed with powerful machine-learning-based classifiers. However, this appearance-based strategy can be implemented in numerous ways and involves many practical considerations. Therefore, an evaluation of popular methods for eye closeness detection is needed.

The major contribution of this paper is to make an extensive comparative study on this approach from the engineering point of view. In particular, we investigated in-depth the influence of several crucial components of an eye closeness detection system, including eye patch alignment, feature extraction and classifiers. Four types of representative feature sets including gray-values, Gabor wavelets, Local Binary Patterns (LBP,[12]), Histograms of Oriented Gradient (HOG,[13]) are compared with respect to three types of classifiers, i.e., Nearest Neighbor (NN), Support Vector Machine (SVM) and AdaBoost on a large benchmark dataset. Both LBP and Gabor feature sets have been used for eye representation before [14][15][16] but HOG not. However, since these features represent different characteristics (local texture, global shape and local shape, respectively) of eyes, it is beneficial to use all of them. Indeed, our experimental results show that fusing various feature descriptors significantly improves the performance of eye closeness detection. In addition, we also show that properly eye patch alignment is important for the performance. These provide useful lessons for the follow-up investigations on this interesting topic.

2   Overall Architecture of Our System

The overall pipeline of our eye closeness detection system is given in Fig. 1. For a given test image, we first detect and crop the face portion using the Viola and Jones face detector [17], then the eyes are localized using the method introduced in [18]. After this, we crop the eye region and align it with those in the training set with method of [19]. On the aligned eye patch various feature sets are extracted and input into the classifier for final decision.

One of the key components of our system lies in the inclusion of eye patch alignment module. This is based on the observation that eyes in a face image may undergo various in-plane/out-of-plane pose changes, or scale changes. Although some feature descriptors are not sensitive to these, other feature sets (e.g., LBP, Gabor wavelets, and HOG) don't have built-in mechanism to handle such variations. Therefore, performing geometric eye normalization is necessary for these descriptors. However, one difficulty for this is that it is hard to find anchor points for eye patches and hence traditional anchor-points-based alignment method can not be applied. Here we adopt an information-theory geometric normalization method originally proposed for medical image registration, i.e., the congealing

method [19]. This is an unsupervised image normalization method which learns a particular affine transformation for each image such that the entropy of a group of eye images is minimized. Fig. 2 gives some illustrations of eye patches normalized using this method, from which we can see that the locations of eyes are centered and their sizes are scaled.
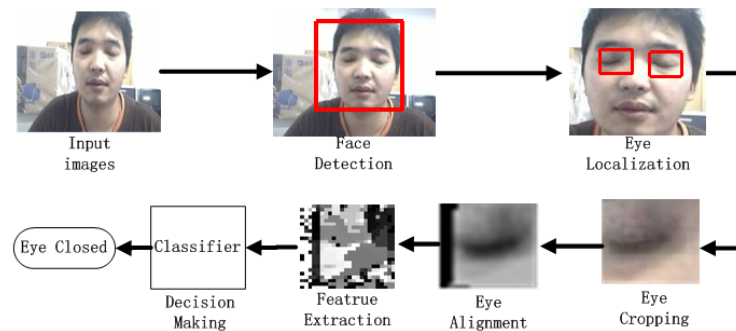

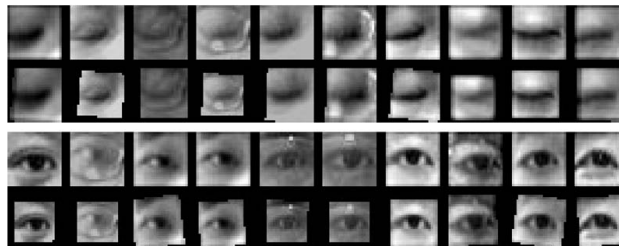**Fig. 1.** The overall architecture of our eye closeness detection system.

Fig. 2. Illustration of eye patches normalized with the congealing method, where patches in the top two rows are original images of closed eyes and their corresponding normalized versions respectively, and patches in the bottom two rows are original images of open eyes and their corresponding normalized images, respectively.

3 Feature Sets

Four types of features are used in the experiments, they are gray-value features, HOG features, Gabor features and LBP feature respectively. The gray feature of an $M \times N$ image patch is simply a $MN \times 1$ column vector. In what follows we briefly describe the LBP feature, Gabor wavelets and the HOG feature.

LBP Feature Local Binary Pattern (LBP) proposed by Ojala [12] is widely used feature descriptor for local image texture. The LBP descriptor has achieved considerable success in various applications such as face recognition and texture recognition, due to its capability to efficiently encode local statistics and geometric characteristics (e.g., spot, flat area, edge and corner) among neighborhood

pixels and its robustness against noise by picking up only 'uniform patterns' for feature description. In this paper, we partition each 24×24 eye patch into 6×6 blocks and represent each block as a 59-dimensional histogram. Therefore for each eye patch we have a 944-dimensional LBP vector (16×59).

Gabor Wavelets Gabor wavelets were originally developed to model the receptive fields of simple cells in the visual cortex and in practice they capture a number of salient visual properties including spatial localization, orientation selectivity and spatial frequency selectivity quite well. They have been widely used in face recognition. Computationally, they are the result of convolving the image with a bank of Gabor filters of different scales and orientations and taking the 'energy image' (pixelwise complex modulus) of each resulting output image. The most commonly used filters in face recognition have the form,

$$\Psi_{\mu,\nu}(z) = \frac{\|k_{\mu,\nu}\|^2}{\sigma^2} * exp(-\frac{\|k_{\mu,\nu}\|^2 * \|z\|^2}{2\sigma^2}) * [exp(ik_{\mu,\nu}, * z - exp(-\frac{\sigma^2}{2}))] \quad (1)$$

where μ and ν define the orientation and scale of the Gabor kernels, $z = (x, y)$, $\|\|$denotes the norm operator, and the wave vector $k_{\mu,\nu}$ is defined:$k_{\mu,\nu} = k_\nu e^{i\varphi_\mu}$, where $k_\nu = k_{max}/f^\nu$ and $\varphi_\mu = \mu\pi/8$. $k_{max}$is the maximum frequency, and f is the spacing factor between kernels in the frequency domain [20]. We use 40 filters with eight orientations and five scales on 24×24 eye patch, then down-sample the resulting vector by 16 to a 1440-dimensional vector.

HOG Feature The aim of Histogram of Oriented Gradients (HOG) feature proposed in [13] is to describe local object appearance and shape within an image by the distribution of intensity gradients or edge directions. To implement these descriptors, we first divide the eye patches into small connected cells of 4×4 in size, and for each cell a histogram (with 9 bins) of gradient directions is calculated, which is then undergone a contrast-normalization within each block, leading to better invariance to changes in illumination or shadowing. Through these steps, we have a 900-dimensional histogram for each patch.

The three aforementioned feature representations of an open eye and a close eye are illustrated in Fig. 3, respectively. One may notice that the LBP feature is good at characterizing detailed texture information of the image, and Gabor wavelets highlight the differences of eye images with different states with respect to global spatial frequency, while the local shape information is best described by the HOG feature.

4   The Classifiers

In this work, we use the Nearest Neighbor, SVM and AdaBoost as our classifiers. The nearest neighbor method is a simple and effective non-parametric classification method and used in this paper as our baseline.

Support Vector Machines Support Vector Machines (SVMs) is the state-of-the-art large margin classifier which has gained popularity within visual pattern recognition. One problem we should handle is imbalance problem. That is, the
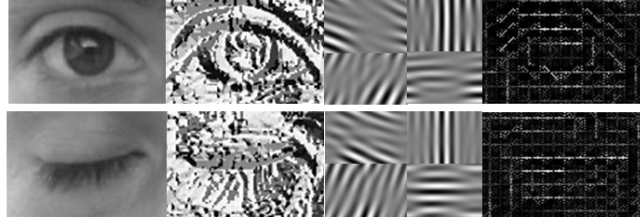
Fig. 3. Illustration of original eye images, their LBP, Gabor wavelet and HOG feature representation, respectively (top row - open eye, bottom row - closed eye). Note that this is only for illustration purpose and the sizes of the eye images are different from those used in the experiments.

number of images of closed eyes and open eyes is different [1], which tends to increase the bias of trained SVM classifier to the class with more samples. To overcome this, before training, we set the penalty respective coefficients for the positive and negative samples to be $\omega_1 = \frac{N^+ + N^-}{2N^+}$, $\omega_2 = \frac{N^+ + N^-}{2N^-}$, where N^+ is the number of positive samples and N^- is the number of negative samples. We used the LIBSVM package [21] with RBF kernel for the SVM-related experiments.

AdaBoost with Pixel-Comparisons As a final classifier compared, we use the AdaBoost, it provides a simple yet effective approach for stagewise learning of a nonlinear classification function. In this work, we use the "difference of intensities of pixels" proposed in [22] as our features. More specifically, we used five types of pixel comparison operators (and their inverses) [23]:

1) $pixel_i > pixel_j$;

2) $pixel_i$ within 5 units (out of 255) of $pixel_j$;

3) $pixel_i$ within 10 units (out of 255) of $pixel_j$;

4) $pixel_i$ within 25 units (out of 255) of $pixel_j$;

5) $pixel_i$ within 50 units (out of 255) of $pixel_j$;

The binary result of each comparison, which is represented numerically as 1 or 0, is used as features. Thus, for an image of 24×24 pixels, there are $2 * 5 * (24 * 24)(24 * (24 - 1))$ or 3312000 pixel-comparison features.

To handle the large number of features, we use Adaboost for feature selection while learning a strong classifier. This is done by mapping each feature to a weak classifier and then selecting the most discriminative weak classifier increasingly for an additive strong classifier at the same time. For more details, see [22]. In our experiments, 2000 weak classifiers are trained and some of them are randomly selected for evaluation in each iteration. We examined the performance achieved from evaluating 1%, 10% and 100% of all possible weak classifiers per iteration.

---

[1] In practice, it is much easier to collect images of open eyes than those of closed eyes.

5   Experiments

5.1   Data and Settings

The data for our experiments are collected from Zhejiang University blinking video database [24]. There are a total of 80 video clips in the blinking video database from 20 individuals, four clips for each individual: one clip for frontal view without glasses, one clip with frontal view and wearing thin rim glasses, one clip for frontal view and black frame glasses, and the last clip with upward view without glasses. We manually select images in each blinking process, including eye images of open, half open, closed, half open, open. In addition, images of the left and the right eyes are collected separately. Some samples of the dataset are shown in Fig. 4. We can see that these images are blurred, with low resolution and may be occluded by glasses.
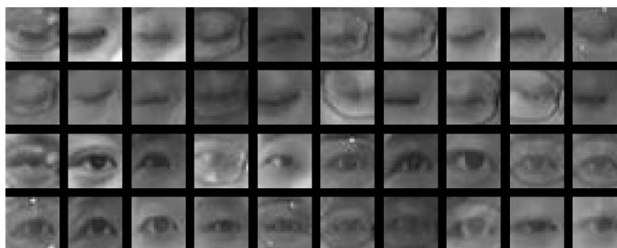

Fig. 4. Illustration of some positive (top two rows) and negative (bottom two rows) samples used for training.

The collected eye images are then divided into two separate sets for training and test purpose. The training set consists of images from the first 16 individuals. The test set consists of the images from the remaining 4 subjects. Note that there is no overlapping in images of subjects between the training set and test set. To further increase the diversity of training samples, various transformations such as rotation, blurring, contrast modification and addition of Gaussian white noise are applied to the initial set of training images, yielding about 6,600 new images in total. Finally, the training set contains 7360 eye images in all, with 1590 closed eye images and 5770 open eye images respectively. The test set is constructed with 410 closed eyes and 1230 open eyes. All these images are geometrically normalized into images of 24×24 pixels.

5.2   Experimental Results

Fig. 5 (left) gives the overall performance of SVM and AdaBoost on our datasets in terms of Receiver Operating Characteristic (ROC) curves. From the results, we can see that the best performer is the SVM using HOG features. In particular, the AUC value of "HOG+SVM" achieves 97.7%, compared to 97.0% for AdaBoost

with 2000 weak classifiers, while the next best performer is the "LBP+SVM". We also see that the LBP features are better than Gabor features for the task of eye closeness detection - their EER values are 5.13 and 7.49 respectively. One possible reason is that the LBP features give a detailed account on the appearance of eye regions while being insensitive to the lighting changes. Actually, when one screws up his eyes, it is really very difficult to make a decision on whether his eyes are closed or not. In these cases, the global shape feature represented by Gabor features is less discriminative than local texture/shape information characterized by the LBP/HOG features.
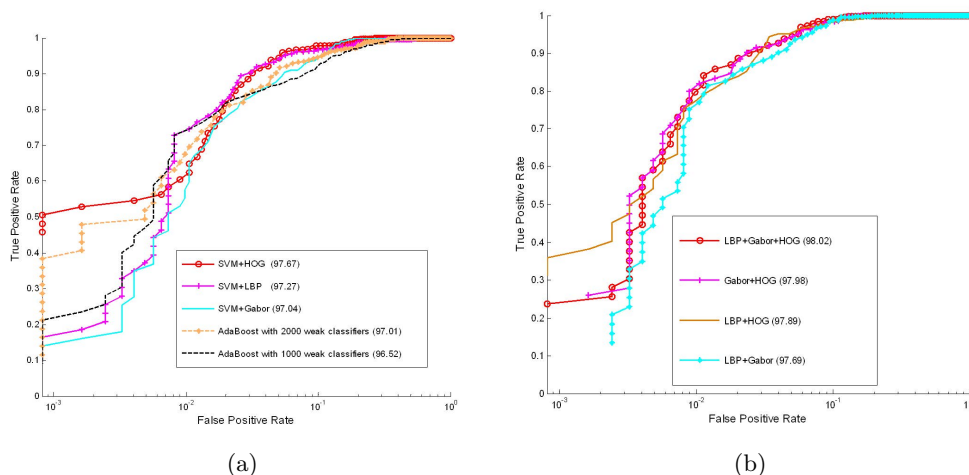

(a)   (b)

Fig. 5. ROC curves of various feature and classifier (left), and of various feature fusion strategies using the SVM classifier (right). AUC values are given in the end of corresponding legend texts.

Table 1 gives the overall comparative performance of various methods and feature sets (the recognition rate is obtained by fixing a threshold learned from the training set). The various nearest neighbor classifier-based schemes serves as the baseline. Again we see that "HOG+SVM" performs best in AUC value but the "LBP+SVM" wins in recognition rate. However, in terms of testing efficiency, the AdaBoost classifier performs faster by at least four times over SVM but with slightly worse ROC performance. This suggests that Adaboost with difference of pixel features is a very attractive candidate in practice due to the high efficiency it provides in testing.

Table 2 and Fig. 5 (right) give the comparative performance using various feature combination schemes. Since LBP, Gabor and HOG feature characterize different aspects of eye patches (i.e, local texture, global shape, and local shape, respectively), it would be useful to fuse the information from three of them. Feature combination could be performed either at the feature level or at the score level and our previous experience shows that score level fusion is simpler to implement (no need to handle the high dimensional problem due to feature concatenation) and usually leads to better results [25]. Hence we adopt the strategy

Table 1. Comparative performance of various features and classifiers

Approach		Recognition Rate(%)	Time(m sec)	AUC(%)	EER(%)
	Gray	73.50	220.25	-	-
	LBP	91.21	195.97	-	-
NN	Gabor	85.65	297.31	-	-
	HOG	91.88	268.62	-	-
	Gray	75.03	12.97	50.00	99.6
	LBP	**95.42**	14.65	97.27	5.13
SVM	Gabor	91.45	12.77	97.04	7.49
	HOG	91.88	19.24	**97.67**	**5.05**
	1%	92.06	**2.890**	96.48	8.30
AdaBoost(1000)	10%	92.91	2.912	96.52	8.71
	100%	92.31	2.980	96.50	8.45
	1%	92.37	2.943	96.67	7.65
AdaBoost(2000)	10%	92.67	3.059	96.73	7.48
	100%	93.47	3.132	97.01	7.08

of score level fusion with z-score normalization in this work. We can see from the table that combining all of the three feature sets gives the best performance both in terms of ROC curve and in terms of recognition rate. It is worthy noting that although the performance of the widely used HOG feature is inferior to the LBP feature in terms of recognition rate, its EER score is higher than LBP, which indicates that the local shape features play an important role in the detection of eye closeness alone. Furthermore, combining the local shape features with local texture descriptor (LBP) significantly improves its performance.

Table 2. Comparative performance of various feature combination schemes (with the SVM classifier)

Approach	Recognition Rate(%)	AUC(%)	EER(%)
LBP	95.42	97.27	5.13
Gabor	91.45	97.04	7.49
HOG	91.88	96.67	5.05
LBP+Gabor	94.69	97.69	5.94
LBP+HOG	95.05	97.89	**4.80**
Gabor+HOG	94.81	97.98	5.05
LBP+Gabor+HOG	**95.42**	**98.02**	5.05

Fig. 6(a) illustrates some of images which are correctly identified as closed eyes using the "HOG/LBP/Gabor+SVM" scheme. Notice that there is large amount of diversity exhibited in the appearance of these closed-eye images. Fig. 6(b) shows some false negative images (upper two rows) and some false positive images (bottom two rows). By carefully examining these images, we can see that they look even confusing to human beings when deciding whether these eyes are open or not. This helps us understand the challenges of eye closeness detection in the real world.
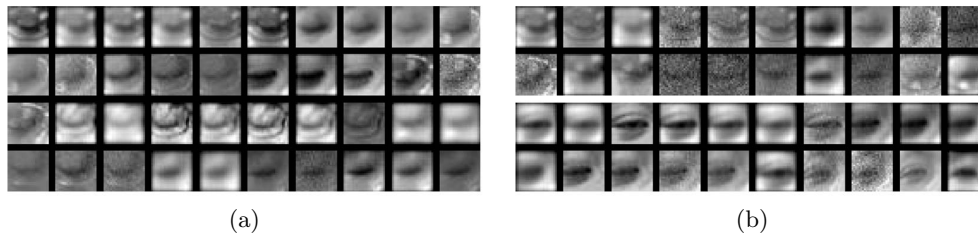
(a)   (b)

Fig. 6. (a)Illustration of images which are correctly identified as closed eyes (true positive), and (b)(upper two rows) images of closed eyes failed to be recognized (false negative) and (bottom two rows) images of open eyes incorrectly recognized as closed eyes (false positive). All the results are with "HOG + LBP + Gabor and SVM" scheme.

Finally, we investigate the influence of eye patch alignment on the performance of the system. As Table 3 shows, adding the module of alignment improves the performance consistently over all the feature sets tested. Although the LBP feature is known to be rotation-invariant and the HOG feature is robust against slight perturbation in the image, they are not robust to general affine transformations. Indeed, Table 3 shows that it is beneficial to do geometric normalization for eye patches before extracting features from them.

Table 3. Comparison of recognition rate w/o eye alignment with SVM.

Processing	LBP	Gabor	HOG
Without alignment	92.7	89.2	89.5
With alignment	**95.4**	**91.5**	**91.9**

6   Conclusions

In this paper, we systematically evaluate several feature sets and classifiers for the task of eye closeness detection. Our experimental results indicate that fusing various feature descriptors significantly improves the performance of the detection system while the AdaBoost classifier with pixel comparisons is a good candidate scheme in practice due to its high efficiency and satisfying performance. In addition, our results show that eye alignment is important and influences the detection accuracy greatly.

Acknowledgements

References

1. Noor, H., Ibrahim, R.: A framework for measurement of humans fatigue level using 2 factors. In: International Conference on Computer and Communication Engineering. (2008) 414–418

2. Eriksson, M., Papanikotopoulos, N.: Eye-tracking for detection of driver fatigue. In: IEEE Conference on Intelligent Transportation System (TSC). (1997) 314–319
3. Mitelman, R., Joshua, M., Adler, A., Bergman, H.: A noninvasive, fast and inexpensive tool for the detection of eye open/closed state in primates. Journal of neuroscience methods **178** (2009) 350–356
4. Sun, R., Ma, Z.: Robust and efficient eye location and its state detection. Advances in Computation and Intelligence (2009) 318–326
5. Valenti, R., Gevers, T.: Accurate eye center location and tracking using isophote curvature. In: Computer Vision and Pattern Recognition (CVPR). (2008) 1–8
6. Wang, H., Zhou, L., Ying, Y.: A novel approach for real time eye state detection in fatigue awareness system. In: Robotics Automation and Mechatronics (RAM). (2010) 528–532
7. Jiao, F., He, G.: Real-time eye detection and tracking under various light conditions. Data Science Journal **6** (2007) 636–640
8. Orozco, J., Roca, F., Gonzàlez, J.: Real-time gaze tracking with appearance-based models. Machine Vision and Applications **20** (2009) 353–364
9. Li, S., Chu, R., Liao, S., Zhang, L.: Illumination invariant face recognition using near-infrared images. IEEE Transactions on Pattern Analysis and Machine Intelligence **29** (2007) 627–639
10. Liu, Z., Ai, H.: Automatic eye state recognition and closed-eye photo correction. Pattern Recognition (2008) 1–4
11. Dehnavi, M., Eshghi, M.: Design and implementation of a real time and train less eye state recognition system. EURASIP Journal on Advances in Signal Processing **2012** (2012) 30
12. Ojala, T., Pietikainen, M., Maenpaa, T.: Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. IEEE Transactions on Pattern Analysis and Machine Intelligence **24**(7) (2002) 971–987
13. Dalal, N., Triggs, B.: Histograms of oriented gradients for human detection. In: Computer Vision and Pattern Recognition (CVPR). (2005) 886–893
14. Cheng, E., Kong, B., Hu, R., Zheng, F.: Eye state detection in facial image based on linear prediction error of wavelet coefficients. In: IEEE International Conference on Robotics and Biomimetics (ROBIO). (2009) 1388–1392
15. Zhou, L., Wang, H.: Open/closed eye recognition by local binary increasing intensity patterns. In: Robotics Automation and Mechatronics (RAM). (2011) 7–11
16. Wang, Q., Yang, J.: Eye location and eye state detection in facial images with unconstrained background. J. Info. & Comp. Science **1** (2006) 284–289
17. Viola, P., Jones, M.: Robust real-time face detection. International Journal of Computer Vision **57** (2007) 137–154
18. Tan, X., Song, F., Zhou, Z., Chen, S.: Enhanced pictorial structures for precise eye localization under incontrolled conditions. In: Computer Vision and Pattern Recognition (CVPR). (2009) 1621–1628
19. Huang, G., Jain, V., Learned-Miller, E.: Unsupervised joint alignment of complex images. In: International Conference on Computer Vision (ICCV). (2007) 1–8
20. Lades, M., Vorbruggen, J., Buhmann, J., Lange, J., von der Malsburg, C., Wurtz, R., Konen, W.: Distortion invariant object recognition in the dynamic link architecture. IEEE Transactions on Computers **42**(3) (1993) 300–311
21. Chang, C., Lin, C.: Libsvm: a library for support vector machines. ACM Transactions on Intelligent Systems and Technology **2** (2011) 27
22. Baluja, S., Sahami, M., Rowley, H.: Efficient face orientation discrimination. In: International Conference on Image Processing (ICIP). (2004) 589–592

23. Baluja, S., Rowley, H.: Boosting sex identification performance. International Journal of Computer Vision **71** (2007) 111–119
24. Database, Z.E.: http://www.cs.zju.edu.cn/ gpan
25. Tan, X., Triggs, B.: Fusing gabor and lbp feature sets for kernel-based face recognition. In: Proceedings of the 3rd international conference on Analysis and modeling of faces and gestures, Springer-Verlag (2007) 235–249