FACE IMAGE NORMALIZATION AND EXPRESSION/POSE VALIDATION FOR THE ANALYSIS OF MACHINE READABLE TRAVEL DOCUMENTS

Markus Storer[1], Martin Urschler[1], Horst Bischof[1], Josef A. Birchbauer[2]

Abstract

Biometrics is a huge and very fast growing domain of methods for uniquely recognizing humans based on one or more intrinsic physical or behavioral traits. Hence, there are many different applications, e.g., surveillance, person verification and identification. The International Civil Aviation Organization (ICAO) provides a number of specifications to prepare automated recognition from travel document photos. The goal of these specifications is increasing security in civil aviation on the basis of standardized biometric data. In this context we are concerned with the strict requirements of deviation from frontal pose, eyes-open, mouth-closed, and eyes looking away for assessing the suitability of face images for inclusion in travel documents. Due to this international standard, there is a high demand for automatically checking face images to assist civil service employees in decision-making. In this work, we present a face normalization and analysis system implementing several parts of the ICAO specification. Our key contribution of this analysis is the fusion of different established classifiers to boost performance of the system. Our results show the superior checking quality of our methods in comparison to state-of-the-art technology of two commercial vendors.

1. Introduction

Analysis of facial images is an important step in biometrics related applications like person verification/identification [1], [17], video surveillance or facial expression assessment [4]. We are specifically interested in checking facial images for their compliance to the International Civil Aviation Organization (ICAO) proposal for machine readable travel documents [7]. The most important parts of the ICAO specification describe a standardized coordinate frame based on face and eye position for potential travel document photographs and a definition of parameters for image quality and facial expression that classify images as suitable or improper for documents like passports. The ICAO specification defines the following rules for accepting photos as suitable according to *deviation from frontal pose*, *eyes-open*, *mouth-closed*, and *eyes looking away* criteria. The full-face frontal pose has to be used. Rotation of the head has to be less than ± 5 degrees from frontal in every direction – up/down, rotated left/right, and tilted left/right. The face expression has to be neutral (non-smiling) with both eyes open normally, looking straight into the camera and the mouth closed. A smile is unacceptable regardless of the inside of the mouth and/or teeth being exposed or not.

[1]Institute for Computer Graphics and Vision, Graz University of Technology, Inffeldgasse 16/II, 8010 Graz, Austria
[2]Siemens IT Solutions and Services, Biometrics Center, Siemens Austria, Strassgangerstrasse 315, 8054 Graz, Austria

Our intended application is an automatic system for checking arbitrary input images for their compliance to the ICAO specification. The ultimate goal is to provide a method that assists civil service employees in determining suitable machine readable travel document photos, thereby increasing efficiency in this selection process and significantly reducing manual work. The main challenge of an automatic system for checking ICAO compliance is its robustness to a large number of different distortions in input images regarding noise, occlusions, bad lighting situations, deviations from frontal pose, and the large variety of human faces with respect to gender, race, or appearance modifications like hair, beard and glasses. In [11], they previously reported results on an automatic face image validation system, where quality aspects of face images are checked.

Face and facial component detection has a long tradition, a survey can be found in [16]. Analysis of facial expressions is a hot topic in recent years [4], [10], with recognition of behavior and emotion from videos as its goal. Face tracking and assessment of medical states like fatigue or bad posture are important, e.g., in driver assistance systems [12] or for the prevention of work-related disorders. In all these applications, deviation from frontal pose is a significant problem [5]. To the best of our knowledge, there is no system reported in the literature, that checks for ICAO compliance, though there are some commercial products.

Our proposed image analysis system consists of a face- and facial component detection stage, that is used to provide input parameters for the face normalization stage. Afterwards, the normalized face image is analyzed according to the ICAO criteria *deviation from frontal pose*, *eyes-open*, *mouth-closed*, and *eyes looking away*. We present a detailed description of these modules in Section 2.. Section 3. presents and discusses the results of applying our analysis system to a large image database. Evaluations are given in the form of relative detection rates compared to commercial products of different vendors. Finally, Section 4. summarizes the work and the main contribution.

Figure 1. Illustration of the overall image analysis system.

2. Face Analysis System Description

Our face analysis system consists of two modules. First, we need to derive a normalized face coordinate system based on eye locations to extract a standardized face image from arbitrary input images. We will refer to this process as *tokenization* according to [7]. The second step takes the tokenized images as input and performs several analysis algorithms in order to derive continuous scores for the likeliness of the corresponding event, e.g., *mouth-closed*. Figure 1 shows the overall workflow of the face analysis system.

2.1. Face Tokenization

The face tokenization module consists of a component for robust face detection from arbitrary images. In our application, we are confronted with input images where only one face is present. Faces are detected using the well known technique AdaBoost [13] incorporating Haar-like- and orientation histogram features. There might be more than one detection from the AdaBoost face detection algorithm [6], so the correct face has to be found in a robust way. In addition, for every potential face detection rectangle, the facial components, mouth and eyes, are located using AdaBoost classifiers [14]. The most likely face is obtained using the face and facial component detections in a normally distributed probabilistic model based on an empirically determined prior distribution of eye and mouth locations. The derived eye locations from the AdaBoost step are further refined by an Active Appearance Model (AAM) [3]. Identified left and right eye coordinates initiate the actual tokenization step according to the ICAO specification [7]. Figure 2 illustrates the tokenization step.

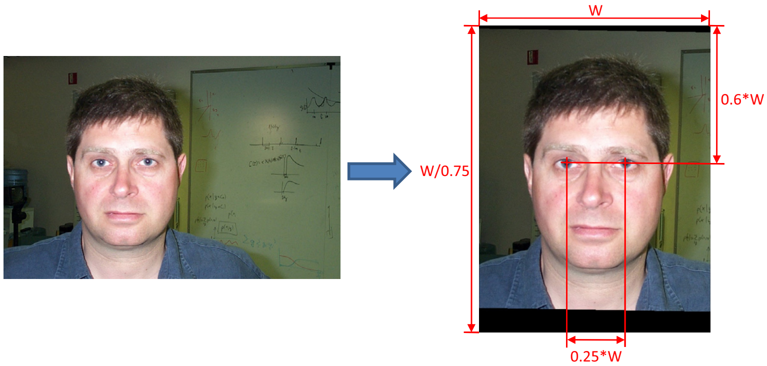

Figure 2. The tokenization procedure. The input image (left) and the tokenized image (right). Image is taken from the Caltech Faces database [2].

2.2. Face Analysis

The face analysis module operates on tokenized images. It performs four classification decisions regarding *deviation from frontal pose*, *eyes-open*, *mouth-closed* and *eyes looking away*. For some of the criteria we make use of the face and facial component detections already obtained from the tokenization step. In order to obtain the necessary robustness, several different classifiers are combined in a fusion step [8]. The performance in terms of detection rate and false-positive rate on a test set is used to derive the weights in the classifier fusion step. The different classifiers used, supplement one another in the fusion step to gain a more robust decision, resulting in lower error rates compared to a single classifier decision. The classifier fusion yields a score value between 0 and 100 for the final decision.

The block diagram of the overall face analysis workflow is illustrated in Figure 3. An analyzed face image is shown in Figure 4. In the following we present the four classification steps in more detail and assume the face component detection has already been performed as described in the previous section.

2.2.1. Analysis of the Deviation from Frontal Pose

The analysis of the *deviation from frontal pose* is carried out using a combination of six different AdaBoost classifiers (Figure 6), that make a decision on frontal-, left-, right-, up- or down pose deviation. There is one classifier which was trained on upwards-looking versus frontal pose face images, another one trained on downwards-looking versus frontal face images. For the deviations from left- and right pose we combined four different classifiers. The first classifier was trained using centered cropped face images, the second classifier took face images, which were cropped taking a shifted window around the face center. In order to incorporate symmetry information in the third classifier, the original image and its mirrored image are overlayed and the average pixel value was taken to form a new image, see Figure 5. This has the advantage, that *deviations from frontal pose*

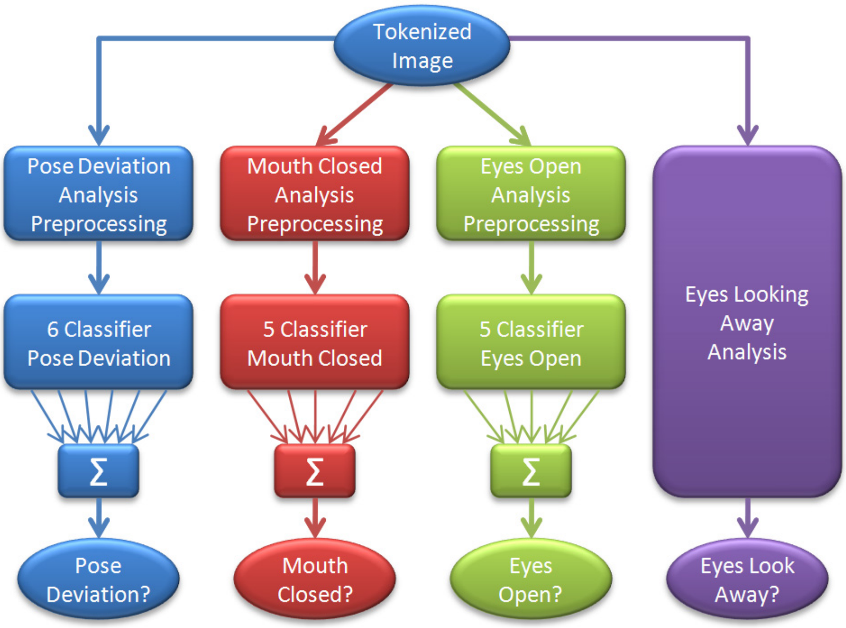

Figure 3. Face analyzer workflow. The classifier fusion for the analysis of the *deviation from frontal pose* is shown in more detail in Figure 6.

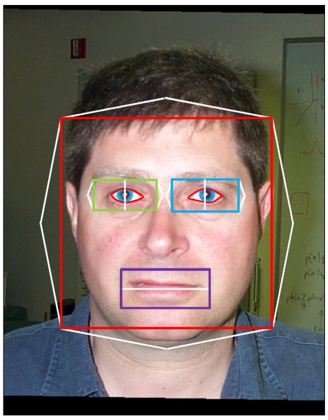

Figure 4. Face analyzer output image. The face rectangle with its four triangles denotes frontal pose. The mouth and eye detections are shown; the lines in these boxes indicate *mouth-closed* (horizontal line) and *eyes-open* (vertical line), respectively. The triangles around the eye detections denote straight gaze. The red contour around the eyes are produced by the AAM and the blue circles show the output of the iris finder.

are "amplified". So the boosting approach is able to discriminate better between pose- and frontal images. This classifier can only determine, either if the face is frontal or not. Hence, a fourth classifier is needed to decide between left- or right pose (see Figure 6). Table 1 gives a summary of the weights.

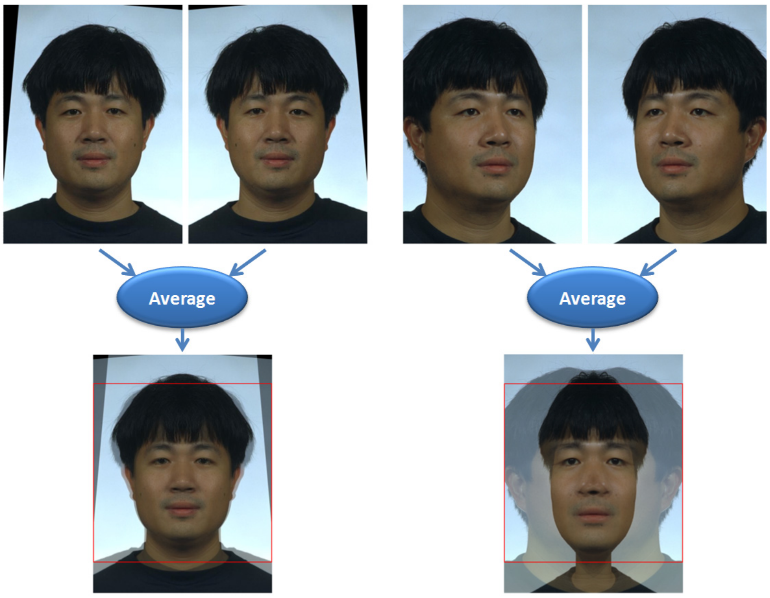

Figure 5. The original image and its mirrored image were overlayed and the average pixel value was taken to form a new image (left figure: frontal image, right figure: image showing pose deviation). The red squares illustrate the cropping area for AdaBoost training.

Table 1. Weights for the classifier fusion of the solely AdaBoost based *deviation from frontal pose* decision.

Classifier	Weight
Left-Centered	39
Left-Shifted	37
Averaged + Left-Right	24

2.2.2. Eyes-Open Analysis

The *eyes-open* classification consists of three different classifier schemes. These are:

- Two different *eyes-open* AdaBoost classifiers (were trained using different feature pools) and one *eyes-closed* AdaBoost classifier.
- An Active Appearance Model [3] for the eyes region of a face.
- Iris localization by a symmetric point of interest detection [9] with an additional Hough voting scheme for radius estimation.

We weight the input of each classifier to the final decision using factors that resemble the performance of the classifiers on the validation data set. See Table 2 for a summary of the weights. Since we get independent decisions for left and right eye, but we only need a single decision for a face image, we take the minimum of the two score values as our final value.

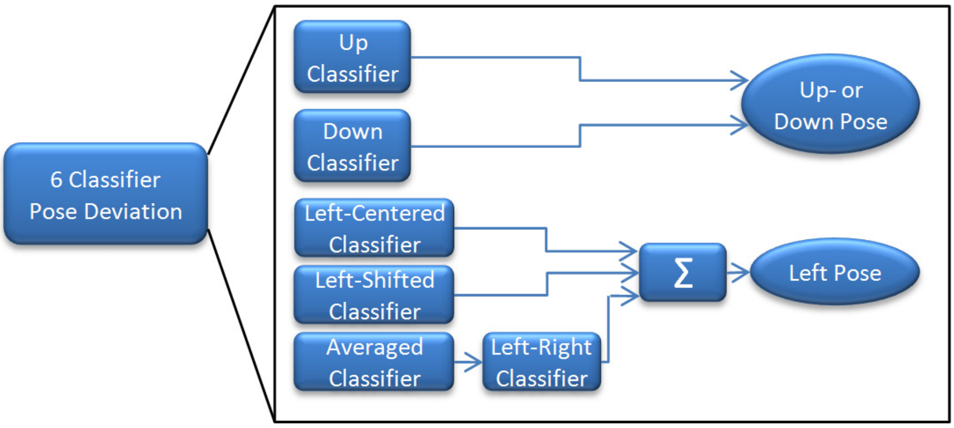

Figure 6. Pose classification workflow.

Table 2. Weights for the classifier fusion of the *eyes-open* decision.

Classifier	Weight
AdaBoost *eyes-open* 1	26
AdaBoost *eyes-open* 2	21
AdaBoost *eyes-closed*	10
AAM	24
Iris detector	19

2.2.3. Mouth Closed Analysis

For the *mouth-closed* analysis, five classifiers in three different domains were chosen:

- An EigenMouth model of closed mouth patches using Principal Component Analysis.
- A simple dark blob detection algorithm that locates dark shadow regions occurring due to an open mouth. In our blob detection algorithm, we use the mouth detection rectangle and enlarge it by ten percent. This input image is transformed into HSV color space and the value coordinate is taken for further processing. After applying an empirically determined threshold of 64 on the mouth patch, we use the blob detection library [6] to extract dark blobs that correspond to shadow regions in the mouth cavity due to an open mouth. We analyze the blob center locations and the blob's compactness. All of these blob analysis methods are used like filters on a set of blobs. If one dark blob region survives all of these filter stages, a *mouth-open* decision is made.
- Two *mouth-closed* AdaBoost classifiers (were trained using different feature pools) and one *mouth-open* classifier.

Classifier fusion is performed as shown in section 2.2.2.. See Table 3 for a summary of the weights of every single classifier to form the final ”strong” classifier.

2.2.4. Eyes Looking Away Analysis

Our strategy uses the response of Gabor filters for determination of the *eyes looking away* criterion. The idea is based on [15]. For this purpose we use the center of the AdaBoost eye detection rectangles

Table 3. **Weights for the classifier fusion of the *mouth-closed* decision.**

Classifier	Weight
EigenMouth (PCA)	10
Blob detector	23
AdaBoost *mouth-closed* 1	21
AdaBoost *mouth-closed* 2	25
AdaBoost *mouth-open*	21

to place three different Gabor filter kernels. The convolution of these Gabor kernels with the input image transformed to value coordinates of the HSV color space at the eye locations are used to determine the strongest and second-strongest response. *Eyes looking away* is performed independently for left and right eye. The continuous score is calculated as a relative measure comparing strongest and second-strongest response.

3. Experimental Results

The performance of the face analyzer was evaluated on our own database. This database consists of 1355 JPEG-compressed face images of 35 individuals. Various scenarios, e.g., *mouth open/closed, eyes open/closed, frontal/left/right pose*, different eye gaze directions, partial face occlusion, are covered equally throughout the database. Example images of this database are given in Figure 7. Note that this database was different from the data that we used for determining the weights in our multi-classifier fusion steps of Section 2. and different from the dataset we used for the training of the AdaBoost classifiers.

Our results are compared to the state-of-the-art technology of two commercial vendors, which we are unfortunately not allowed to mention, therefore we refer to them as vendor ’X’ and vendor ’Y’. Figure 8, Figure 9, and Figure 10 show the comparisons of the relative detection rate for the analysis of the *deviation from frontal pose*, *eyes-open* analysis, and *mouth-closed* analysis, respectively. For each classifier, we show our results using the best performing system as the reference to which the other systems are compared. This relative performance comparison is presented at several different false acceptance rate (FAR) configurations. The comparisons show the good performance of our different face analysis modules with respect to the other vendors. For all three evaluations we are either in the same range or even outperforming the other systems.

Another aspect that we learned from our experiments is the superior performance of the multi-classifier methods compared to the results using only the individual classifiers. By choosing to combine several kinds of methods complementary to each other we are able to increase the classification result and make the method more robust, which results in a reduction of the equal-error-rate (EER) of about 6 up to 15 percent. We omit the full comparison of single versus multi-classifier methods due to lack of space. However, we will present these results in a future publication.

We also performed evaluation experiments for *eyes looking away*. Our final version reaches a correct detection rate of around 80 percent. This moderate performance is due to inaccuracies in determining eye positions, because in our database, there are many persons wearing glasses. Here we have no comparisons, because there are no implementations of *eyes looking away* from different vendors so far.

Figure 7. Sample images from the database for evaluation: (a) frontal pose, eyes open, mouth closed, (b) left pose, eyes open, mouth closed, (c) right pose, eyes open, mouth closed, (d) frontal pose, eyes closed, mouth closed, (e) frontal pose, eyes open, mouth open, (e) frontal pose, eyes open, mouth open.

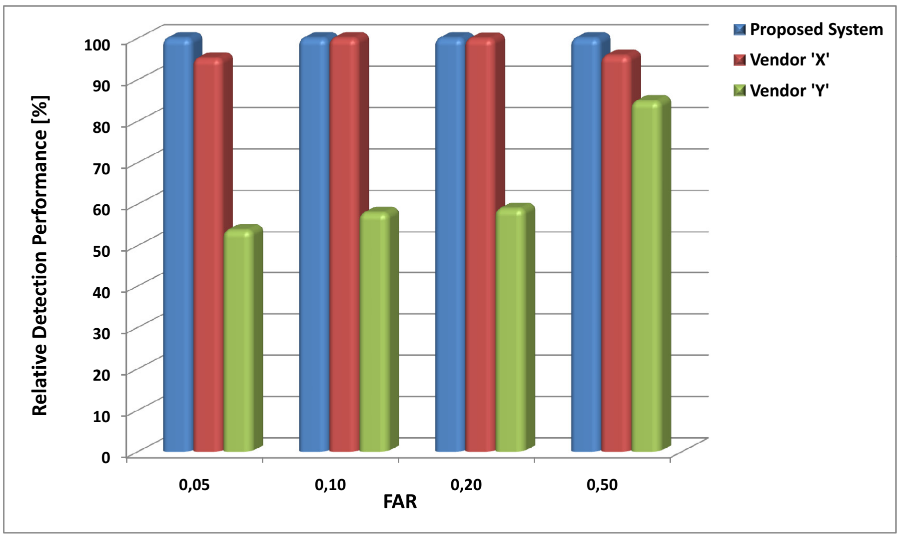

Figure 8. Comparison of the relative detection rates for the analysis of the *deviation from frontal pose*.

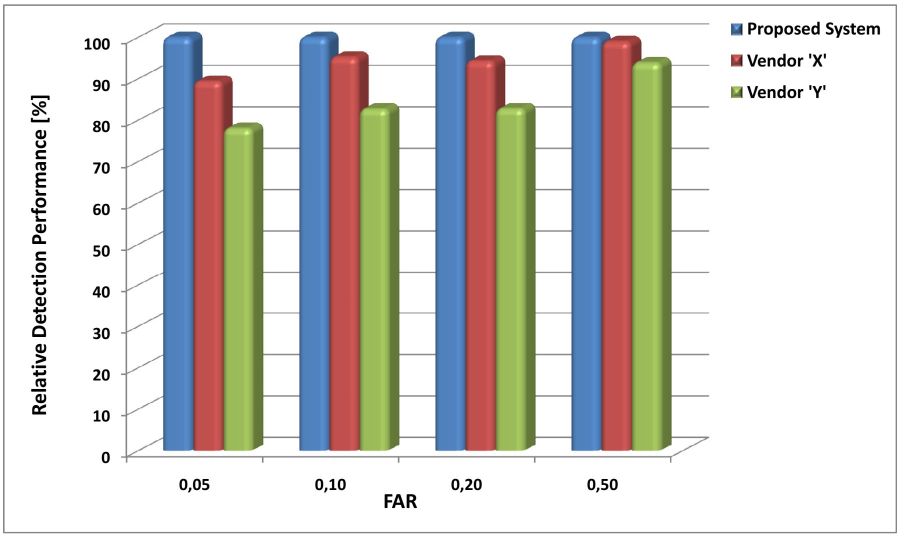

Figure 9. Comparison of the relative detection rates for the *eyes-open* analysis.

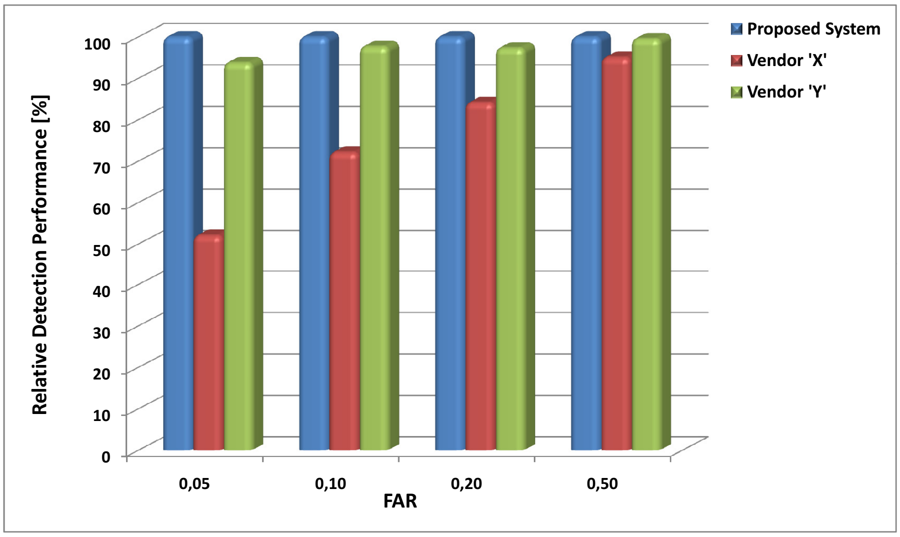

Figure 10. Comparison of the relative detection rates for the *mouth-closed* analysis.

4. Conclusion

This work presents our face normalization and analysis system for checking ICAO compliance. In the system, arbitrary face images have to be tokenized in a first step to bring them into a format compliant to the ICAO specifications. In a second step the strict criteria, how an internationally complying passport photo has to look like, are met by analyzing tokenized images for *deviation from frontal pose*, *eyes-open*, *mouth-closed*, and *eyes looking away*. Therefore, different classifiers are fused to form a final "strong" classifier for checking all criteria. This fusion gives a great boost in the analysis performance compared to the use of individual classifiers. Results show the competitive to superior quality of facial checks in comparison to state-of-the-art technology of two commercial vendors.

Further work is necessary to still improve the performance of the *eyes-open* criterion. The presented method has problems with people with glasses and occluded eyes. We will also prepare a more detailed comparison of the single versus multi-classifier methods in a future publication.

Acknowledgement

This work has been funded by the Biometrics Center of Siemens IT Solutions and Services, Siemens Austria.

References

[1] Andrea F. Abate, Michele Nappi, Daniel Riccio, and Gabriele Sabatino. 2D and 3D face recognition: A survey. *Pattern Recognition Letters*, 28(14):1885–1906, January 2007.

[2] Caltech. Caltech Faces database. http://www.vision.caltech.edu/html-files/archive.html, 1999.

[3] T. F. Cootes, G. J. Edwards, and C. J. Taylor. Active appearance models. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 23(6):681–685, 2001.

[4] Beat Fasel and Juergen Luettin. Automatic facial expression analysis: A survey. *Pattern Recognition*, 36(1):259–275, 2003.

[5] Ralph Gross, Simon Baker, Iain Matthews, and Takeo Kanade. Face recognition across pose and illumination. In Stan Z. Li and Anil K. Jain, editors, *Handbook of Face Recognition*. Springer-Verlag, June 2004.

[6] Intel. Open computer vision (OpenCV) library. http://www.intel.com/technology/computing/opencv/index.htm, 2007.

[7] International Civil Aviation Organization (ICAO). Doc 9303, machine readable travel documents, part 1 - machine readable passport, sixth edition, 2006.

[8] Josef Kittler, Mohamad Hatef, Robert P.W. Duin, and Jiri Matas. On combining classifiers. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 20(3):226–239, March 1998.

[9] Gareth Loy and Alexander Zelinsky. Fast radial symmetry for detecting points of interest. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 25(8):959–973, August 2003.

[10] Maja Pantic and Leon J. M. Rothkrantz. Automatic analysis of facial expressions: The state of the art. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(12):1424–1445, December 2000.

[11] M. Subasic, S. Loncaric, T. Petkovic, H. Bogunovic, and V. Krivec. Face image validation system. In *Proceedings 4th International Symposium on Image and Signal Processing and Analysis*, pages 30–33, 2005.

[12] David Tock and Ian Craw. Tracking and measuring drivers' eyes. In *Proceedings 7th British Machine Vision Conference*, volume 14, pages 541–547, August 1996.

[13] Paul Viola and Michael J. Jones. Rapid object detection using a boosted cascade of simple features. In *Proceedings IEEE Conference on Computer Vision and Pattern Recognition*, volume 1, pages 511–518, 2001.

[14] Paul Viola and Michael J. Jones. Robust real-time face detection. *International Journal of Computer Vision*, 57(2):137–154, 2004.

[15] U. Weidenbacher, G. Layher, P. Bayerl, and H. Neumann. Detection of head and gaze direction for human-computer interaction. In *Perception and Interactive Technologies*, number 4021 in LNAI, pages 9–19, 2006.

[16] Ming Hsuan Yang, David Kriegman, and Narendra Ahuja. Detecting faces in images: A survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(1):34–58, January 2002.

[17] W. Zhao, R. Chellappa, P. J. Phillips, and A. Rosenfeld. Face recognition: A literature survey. *ACM Computing Surveys*, 35(4):399–458, 2003.