Classifier Fusion for Robust ICAO Compliant Face Analysis

Markus Storer, Martin Urschler and Horst Bischof
Institute for Computer Graphics and Vision
Graz University of Technology
Inffeldgasse 16/II, 8010 Graz, Austria
{storer, urschler, bischof}@icg.tugraz.at

Josef A. Birchbauer
Siemens Biometrics Center
Siemens IT Solutions and Services
Strassgangerstrasse 315, 8054 Graz, Austria
josef-alois.birchbauer@siemens.com

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 with applications in many different areas, 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 to increase security in civil aviation on the basis of standardized biometric data. 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 overall system. Our results show the superior checking quality on facial images due to utilizing classifier fusion compared to a single classifier decision.

1. Introduction

Biometric related applications like person verification/identification [2], [28], [11], video surveillance or facial expression assessment [5] require the analysis of facial images. Of specific interest in security related applications is to check facial images for their compliance to the International Civil Aviation Organization (ICAO) specification for machine readable travel documents [8]. The ICAO specification describes 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, identification cards or visas.

We focus on a part of our automatic system for checking arbitrary input images for their compliance to the ICAO specification. The ultimate goal of the overall system 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 type (or lack of hair), beard, make-up and glasses. All these aspects make it particularly difficult to design a robust system for face analysis. Up to our knowledge there have been no previous publications on this topic besides [20], reporting results on an automatic face image validation system, where a number of rather simple quality aspects of face images are checked. However, a number of commercial products currently exist for ICAO compliant face analysis.

Figure 1. Illustration of the overall image analysis system consisting of normalization (*tokenization*) and compliance analysis.

Face and facial component detection and related analysis has a long tradition in the computer vision literature, a survey can be found in [27]. Analysis of facial expressions is a hot topic in recent years [5], [17], 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 [21] or for the prevention of work-related disorders.

Our specific interest lies in the analysis of the state of the eyes and the mouth on a given arbitrary face image. In our system first a normalization stage (which is called *tokenization* according to [8]) transforms an arbitrary input image to a normalized coordinate frame depending on the eye

978-1-4244-2154-1/08/$25.00

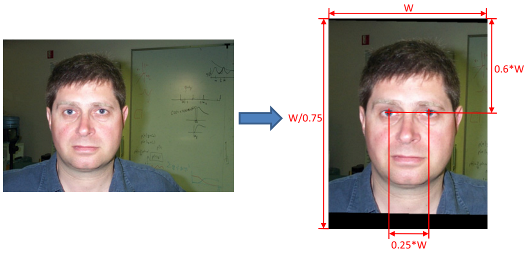

Figure 2. The tokenization procedure. The input image (left) and the tokenized image (right). Image taken from the Caltech Face database [1].

Figure 3. Some sample images from the evaluation database: (a) eyes open, mouth closed, (b) eyes open, mouth open, (c) eyes closed, mouth open. Taken from the AR face database [15].

positions by making use of a robust face and facial component detection algorithm. The overall analysis pipeline is depicted in Figure 1, while the outcome of the *tokenization* process is illustrated in Figure 2.

After the detection of eyes and mouth components in a facial image, an analysis procedure is applied to assess the ICAO criteria *eyes-open* and *mouth-closed*. Reviewing the literature revealed a multitude of techniques being applied to this problem, among them machine-learning approaches like support vector machines [23], or Boosting [19], model-based approaches like EigenFaces [22] or Active Appearance Models [4], or simpler geometric and template based methods for detecting eye- or mouth-related features like lips, teeth, iris, or eyelids. Despite their usefulness in many situations, all of these approaches have their specific drawbacks, e.g., performance of model-based approaches decreases significantly in the presence of outliers, while their accuracy is superior to other methods if the model fitting is successful. From the observation of differing performance of different - to a certain extent complementary - algorithms, we adapted the interpretation of each algorithm as a single expert giving a vote for a certain classification decision. By combining the votes of all classifiers in a classifier fusion scheme [9] , [10], we state the hypothesis that the performance of the combined scheme is superior compared to the performance of the single classifiers in the ensemble. This makes the decision for the specific events *eyes-open* and *mouth-closed* more robust in the presence of difficult situations like noisy input data, lighting conditions or partial occlusions, e.g., wearing glasses.

In the following we will validate our hypothesis by presenting our face analysis system consisting of the single classifiers (Section 2) and the classifier fusion strategies in more detail in Section 3. Section 4 shows the results of our experiments on two databases. Finally we discuss and summarize our findings in Section 5.

2. Face Analysis System Description

Our face analysis system operates on *tokenized* images. It performs several classification decisions of which we restrict ourselves to *eyes-open* and *mouth-closed* events. These criteria rely on our facial component detection stage, where a robust scheme performs face, eye and mouth localization from face component hypotheses in a probabilistic voting framework.

The ICAO specification [8] defines the following rules for accepting photos as suitable according to *eyes-open* and *mouth-closed* criteria. The face expression has to be neutral (non-smiling) with both eyes open normally and the mouth closed (see Figure 3). A smile is unacceptable regardless of the inside of the mouth and/or teeth being exposed or not. Starting from this specification and taking the large variety of possible problems in real-world images (due to noisy data, inappropriate lighting situations, occlusions due to hair or glasses, or the large variety in appearance of different people) it is intuitive that a single classification method will not be able to solve this task in a robust manner. Therefore several different classifiers are combined in a fusion step (see Figure 4). An important assumption for efficiently combining classifiers is that they show complementary behavior and their estimates are as independent as possible. In practice it is very hard to come up with a set of totally independent methods, so one has to rely on experimental evaluation to show their applicability to a given task.

For the training based approaches we have used a large manually annotated training set of around 4600 face images which were taken from the Caltech Face database [1], the FERET database [18] and from a third database constructed from our own images.

2.1. Eyes Open Analysis

For the analysis of the event *eyes-open* we use an ensemble of five classifiers. Two classifiers are based on AdaBoost, one uses the Active Appearance Model , one is derived from the *EigenFace* method and the last method is based on a geometric iris localization strategy. The *eyes-open* decision is performed independently for the left and right eye and leads to a confidence value $d_i(\mathbf{x}) \in [0, 1]$ representing the range between closed and open eyes. The minimum of these two separate decisions forms the final result,

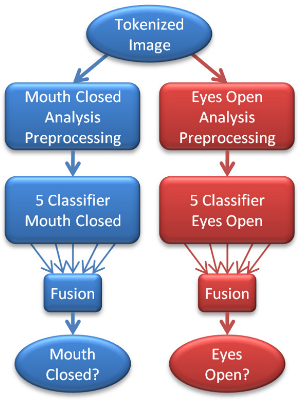

Figure 4. Face analyzer workflow. From tokenized images we perform some pre-processing, apply the single classifiers and fuse their results to form a final decision.

since one closed eye already corresponds to an *eyes-closed* decision.

2.1.1 Active Appearance Model

We trained an Active Appearance Model (AAM) [4], (Appendix B), **(CL1)**, for face image regions around the eyes, see Figure 5. Our training set consists of 427 manually annotated face images taken from the Caltech Face database and our own collection. Training images show variations in the opening of the eyes, slight pose variations, and eyes, which are looking straight and away. For model building we keep 97.5 percent of the eigenvalue energy spectrum to represent our compact model. To apply the AAM to a given image for *eyes-open* classification, we initialize the mean shape of the AAM by the roughly estimated left and right eye locations from facial component detection. To derive a measure of the likelihood of the *eyes-open* event we analyze the vertical eyes' opening of the converged AAM shape model in the left and right eye area respectively. We compare the opening to a pre-defined threshold $T_{E,aam}$ and additionally weight the distance to the threshold with the AAM residual error that represents an estimate of success or failure of model fitting.

2.1.2 Iris Detection Approach

Our geometric iris detection approach **(CL2)** is based on a fast radial symmetry detector presented in [14]. For each eye we restrict ourselves to an image patch around the eye. After performing edge-preserving anisotropic smoothing [25], we calculate a symmetry transform image by estimating gradient orientation and magnitude projection images over several scales according to [14]. Local minima of the symmetry transform image correspond to centers of radial-symmetric structures. The strongest response of this

(a) (b)

Figure 5. Active Appearance Model of the eye region. (a) Learned mean shape/texture and the texture after successful fitting. (b) AAM shape model after successful fitting drawn on the input image.

transform corresponds to iris centers. Afterwards we perform a more accurate iris radius estimation by using a one-dimensional Hough voting on the binary response image from a Canny edge detector [3]. We favor iris radii that are conform with a rough scale estimation of the iris that we are able to derive from our tokenized input images. The voting histogram entry with the maximal response gives the desired iris radius. From the strength of the symmetry image minimum and the voting histogram we derive a confidence measure for the *eyes-open* event.

2.1.3 AdaBoost Classifier

Eyes-open analysis using AdaBoost [19], [24], (Appendix A), utilizes two different classifiers, both trained with the OpenCV [7] library. These classifiers focus on Haar wavelet filter features. The first one **(CL3)** was trained on image patches of closed eyes and the second **(CL4)** on open eyes. For the closed eye classifier 464 positive image patches were used, while the open eye classifier was trained with 2732 image patches. The discrepancy in the number of positives is due to the unequal representation of both classes in our training set. In both cases the set of negative images was taken from generic background images.

Our classification strategy for each trained Boosting classifier takes the approximate location of the eye from the facial component detection stage and applies the classifier to a slightly enlarged region around this region. That is the reason, why we trained one open-eye and one closed-eye detector utilizing the sliding window approach in the enlarged region, hence exploiting the detector as a classifier. If we detect an open eye rectangle we report a confidence measure according to [26]. If we detect a closed eye rectangle we report the inverse of this confidence measure.

2.1.4 EigenEye Model

The *EigenFace* model [22], (Appendix C), is a generative method that explicitly models face images in the face image space. The same principle can be used to model open eyes (*EigenEye*, **CL5**). This generative *EigenEye* model is based on a training set of approximately 2700 open eye images for left and right eye, respectively. These images are coarsely registered before model calculation. From the compact PCA representation we keep 120 eigenvectors, resembling 97.5 percent of the energy in the eigenvalue spectrum. For deciding on the similarity to the space of open eyes we define a threshold $T_{E,pca}$ on the reconstruction error. The distance to this threshold gives a continuous confidence measure.

2.2. Mouth Closed Analysis

For the *mouth-closed* analysis we have also used an ensemble of five classifiers. Three classifiers are based on AdaBoost, one approach makes use of the *EigenFace* framework and the last classifier utilizes a blob detection algorithm that locates dark blobs due to mouth cavity shadows. Decision scores $d_i(\mathbf{x}) \in [0, 1]$ range between 0 for open and 1 for closed mouthes.

2.2.1 Geometric Dark Blob Analysis

The dark blob analysis **(CL1)** is a geometric method that makes use of the fact that open mouthes very often exhibit dark blobs due to shadows in the open mouth cavity compared to the rest of a mouth image patch. Therefore, we investigate a slightly enlarged version of the mouth detection area, transform it into HSV color space and proceed by working solely on the Value coordinate. After binarizing the mouth patch using thresholding [16], we perform a blob detection process that extracts dark blobs corresponding to shadow regions. A filtering stage on the extracted blobs regarding their size, center locations and compactness removes unlikely shadow regions that may occur, e.g., due to beards. If a dark blob region survives this filtering stage we decide for the *mouth-open* event, otherwise for *mouth-closed*. A confidence measure is derived from the size of the detected blob region.

2.2.2 Boosting Classifier

Mouth-closed analysis using AdaBoost (Appendix A) leads to three different classifiers. The first one **(CL2)** is trained with the OpenCV library using 3785 closed mouth patches as positives and a large pool of non-mouth patches as negatives. This classifier focuses on Haar wavelet filter features. The classification strategy takes the approximate location of the mouth from the facial component detection stage into account and applies the classifier to a slightly enlarged region around the mouth.

The second AdaBoost classifier **(CL4)** uses integral image approximations of edge orientation histograms. The weight update strategy follows the *RealBoost* scheme. We expect complementary behavior of our *RealBoost* approach to the OpenCV implementation due to the different features under consideration, i.e., the OpenCV library focuses on wavelet filter approximations, while our *RealBoost* learns features from the edge information of an image. *RealBoost* is applied similar to the first classifier but uses 2475 open mouth patches as positive images in the training stage.

The third AdaBoost classifier **(CL3)** is also trained with the *RealBoost* scheme on 1200 closed mouth patches. This classifier uses the same amount of open mouth patches as negatives and can only be applied directly (without a sliding window approach) to the detected mouth patches from the facial component detection. All of the AdaBoost classifiers report a confidence measure which is calculated according to [26] in case of closed mouths.

2.2.3 EigenMouth Model

Similar to the *EigenEye* model, the generative *EigenMouth* model **(CL5)** is based on a training set of 3785 closed mouth images in neutral expression. These images are coarsely registered and the model is calculated. From the compact PCA representation we keep 140 eigenvectors, resembling 97.5 percent of the energy in the eigenvalue spectrum. For deciding on the similarity to the space of closed mouthes we define a threshold $T_{M,pca}$ on the reconstruction error. The distance to this threshold gives a continuous confidence measure.

3. Classifier Fusion

We hypothesize that fusing multiple classifiers generates more accurate classification results compared to single classifier decisions. Hence, our goal is to evaluate different fusion strategies to combine the classifiers discussed in the previous sections. Let D denote a single classifier and $\mathbf{x} \in \mathbb{R}^n$ a feature vector representing a pattern to be classified. The classifier represents a mapping

$$D: \ \mathbf{x} \in \mathbb{R}^n \rightarrow \omega_j \in \Omega \,,$$

where ω_j is one of the c possible classes of $\Omega = \{\omega_1, \dots, \omega_c\}$. Denote $\{D_1, \dots, D_L\}$ as the set of L classifiers. The output of the ith classifier is $D_i(\mathbf{x}_i) = [d_{i,1}(\mathbf{x}_i), \dots, d_{i,c}(\mathbf{x}_i)]^T$, where $\mathbf{x}_i$ is the specific feature vector representation of the input pattern needed by classifier D_i and $d_{i,j}(\mathbf{x}_i)$ is the confidence, i.e., the degree of support, classifier D_i assigns to the assumption of $\mathbf{x}_i$ originating from class j. The fused output $\hat{D}$ of the L single classifiers is

$$\hat{D}(\mathbf{x}) = \mathcal{F}(D_1(\mathbf{x}), \dots, D_L(\mathbf{x})), \qquad (1)$$

where $\mathcal{F}$ is called the *fusion strategy*. Resulting from $\hat{D}$, the final confidence values assigned to each class are $\hat{d}_j$. The following fusion strategies are investigated:

Majority vote (MAJ) The outputs of the single classifiers $d_{i,j}$ are assigned to a specific class explicitly. The class label that occurs most often is taken as the final decision.

Minimum (MIN) $\hat{d}_j(\mathbf{x}) = \min_i \{d_{i,j}(\mathbf{x})\}$

Maximum (MAX) $\hat{d}_j(\mathbf{x}) = \max_i \{d_{i,j}(\mathbf{x})\}$

Average (AVR) $\hat{d}_j(\mathbf{x}) = \frac{1}{L} \sum_{i=1}^{L} d_{i,j}(\mathbf{x})$

Binarized Average (BAVR) This scheme is equivalent to the average fusion strategy, except for the outputs of the single classifiers $d_{i,j}$ being assigned to a specific class explicitly before averaging.

Product (PRO) $\hat{d}_j(\mathbf{x}) = \prod_{i=1}^{L} d_{i,j}(\mathbf{x})$

Prior Confidence (PRIOR) A priori confidences of the single classifiers are obtained from tests on a validation dataset according to their performance exhibited (Table 1). The prior confidences are accumulated according to the decision of the corresponding single classifier.

Bayes Combination (BAYES) This scheme assumes that the classifiers are mutually independent and that the posterior confidences of the single classifiers are equal to posterior probabilities. For our single classifiers we expect independence, because different underlying concepts and methodologies are used, e.g., different features for classification.

The fusion strategies presented above can only be justified under strict probabilistic conditions. Nevertheless, some of them exhibit excellent performance as can be seen in the experimental results, a fact which was already stated in [9] for some of the classifier fusion strategies. The strict probabilistic conditions motivate to learn the fusion of the single classifiers. Thus, we use a support vector machine (SVM) [23], [12] to learn the optimal fusion in a linear combination.

SVM learned posterior confidence fusion (SVM POST) Obtained posterior confidences on a labeled validation dataset (independent of the training- and test set) are used for the learning of a linear combination of the single classifiers by the use of a SVM.

SVM learned binary posterior confidence fusion (SVM BPOST) Obtained posterior confidences on a validation dataset are assigned to a specific class explicitly. These binarized confidences are then used for learning the optimal combination by the use of a SVM.

SVM learned Bayes combination (SVM BAYES) In this scheme the optimal Bayes combination of the single classifiers confidences is learned by the use of a SVM on a validation dataset.

4. Experimental Results

We used two different face databases for the evaluation of our single classifiers and fusion strategies. The first one is the publicly available **AR face database** [15]. It consists of 126 unique people, frontal view face images (70 male, 56 female) of different facial expressions, illumination conditions and occlusions resulting in a total amount of over 4000 color images of a resolution of 768 by 576. We used all the images except those with dark sunglasses or occluded mouth due to a scarf yielding about 1700 images for evaluation. Annotation data is available on request. Some samples of this challenging database are given in Figure 6. The **second database** we used is a private data set containing 325 frontal face color images of 480 by 640 pixels with 30 people showing different facial expressions.

Figure 6. Sample images from the AR face database [15] showing the difficulties of the images under consideration.

The evaluation procedure is exemplified on our own database for the AAM *eyes-open* classifier in Figure 7. All evaluation results on the AR face database are summarized in Table 2 for *eyes-open* and in Table 4 for *mouth-closed* analysis. The evaluation results on our private database are presented in Table 3 for *eyes-open* and Table 5 for *mouth-closed* analysis. The comparison of the ROC curves of the best single classifier to the best fusion strategy is shown in Figure 8 for the AR face database and in Figure 9 for our own database.

Table 1. Prior confidences of the single classifiers. (left) eyes-open classifiers, (right) mouth-closed classifiers

Classifier	Prior
CL1	0.25
CL2	0.20
CL3	0.25
CL4	0.19
CL5	0.11

Classifier	Prior
CL1	0.23
CL2	0.21
CL3	0.25
CL4	0.21
CL5	0.10

The best overall fusion performance is exhibited by the learning based SVM approaches, and here especially the SVM using the posterior confidences. However, also simple fusion strategies, e.g., average, show very good results. The single classifier performance on our private database is slightly better compared to the AR database due to a larger complexity of the images contained in the latter data set. The AR face database contains a large number of images from people wearing different glasses, often show severe specular reflections, simulate bad illumination conditions and extremely wide open mouths, where we have problems of robustly locating the mouth region. However, the hypothesis of fusing multiple classifiers generating more accurate and robust classification results compared to a single classifier, is thus approved, illustrated in our figures and tables.

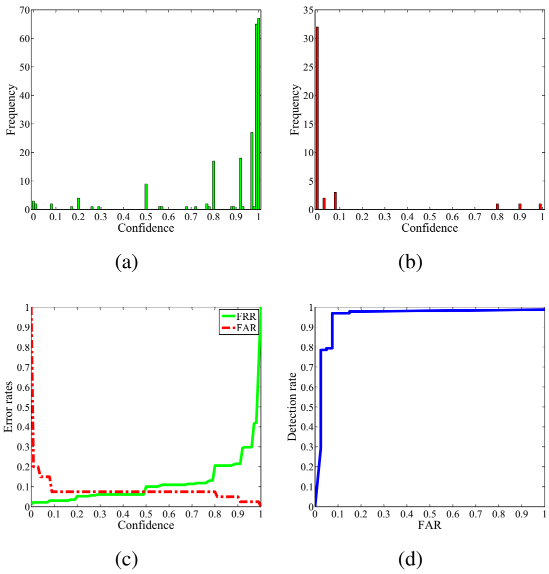

Figure 7. Evaluation procedure for the eyes-open AAM on our own database. (a) Distribution of the confidence values for open-eyes in the database. (b) Distribution of the confidence values for closed-eyes in the database. (c) Characteristic of the false rejection rate (FRR) and false acceptance rate (FAR). (d) ROC curve

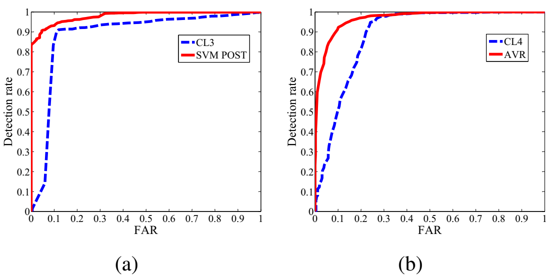

Figure 8. Comparison of the ROC curves of the best single classifier to the best fusion strategy for the (a) eyes-open and (b) mouth-closed analysis on the AR face database.

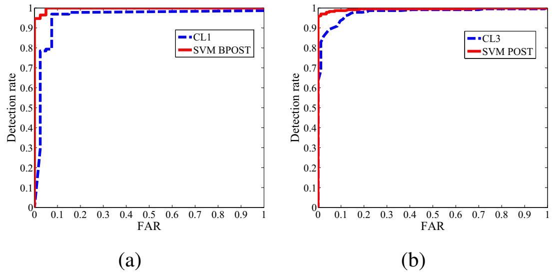

Figure 9. Comparison of the ROC curves of the best single classifier to the best fusion strategy for the (a) eyes-open and (b) mouth-closed analysis on our own face database.

Table 2. Evaluation results for the eyes open analysis on the AR face database

	EER	Detection Rate @ FAR			
		0.05	0.1	0.3	0.5
CL1	18.35	48.90	68.68	86.66	94.14
CL2	29.37	70.63	70.63	70.63	70.63
CL3	10.85	11.96	85.18	93.49	95.04
CL4	24.89	74.45	74.61	75.37	76.65
CL5	25.75	47.87	63.86	76.04	81.12
MAJ	18.08	81.92	81.92	81.92	81.92
MIN	26.91	71.38	71.77	73.34	75.44
MAX	15.91	29.00	62.17	99.13	99.73
AVR	9.21	87.96	91.17	96.23	99.66
BAVR	7.82	81.92	92.18	92.18	99.73
PRO	28.03	70.59	70.89	72.09	73.28
PRIOR	7.82	90.55	92.18	92.18	99.73
BAYES	8.81	89.22	91.65	96.99	99.82
SVM POST	7.65	90.86	93.19	97.97	99.86
SVM BPOST	8.26	85.72	92.11	99.25	99.73
SVM BAYES	11.31	85.18	87.91	96.03	99.86

Table 3. Evaluation results for the eyes open analysis on our own face database

	EER	Detection Rate @ FAR			
		0.05	0.1	0.3	0.5
CL1	7.50	79.39	96.93	97.92	98.14
CL2	10.96	89.04	89.04	89.04	89.04
CL3	13.51	85.09	85.53	92.65	94.88
CL4	12.72	87.28	87.28	87.28	87.28
CL5	15.30	67.11	83.77	88.16	90.94
MAJ	5.70	94.30	94.30	94.30	94.30
MIN	13.16	86.84	86.84	96.38	97.04
MAX	6.50	87.72	99.56	100.00	100.00
AVR	3.95	96.05	96.49	100.00	100.00
BAVR	5.00	96.05	96.05	100.00	100.00
PRO	12.97	86.64	86.89	87.85	88.82
PRIOR	3.95	96.05	99.56	100.00	100.00
BAYES	3.73	96.49	96.49	100.00	100.00
SVM POST	3.51	96.49	98.25	100.00	100.00
SVM BPOST	3.51	100.00	100.00	100.00	100.00
SVM BAYES	5.00	95.18	98.25	100.00	100.00

5. Conclusion

This work presents our face analysis system for checking ICAO specification compliance. ICAO requirements on facial images for machine readable travel documents need to be checked automatically in a variety of applica-

Table 4. Evaluation results for the mouth closed analysis on the AR face database

	EER	Detection Rate @ FAR			
		0.05	0.1	0.3	0.5
CL1	36.90	0.00	0.00	0.00	92.31
CL2	28.33	64.83	66.03	72.18	80.17
CL3	25.13	31.49	44.88	79.67	95.31
CL4	19.53	25.34	50.80	97.83	99.67
CL5	23.90	47.66	60.63	81.36	90.31
MAJ	17.12	24.20	48.39	97.16	97.16
MIN	16.91	60.36	70.61	89.66	89.76
MAX	39.77	42.49	43.16	45.82	98.26
AVR	9.40	81.12	91.90	98.33	99.75
BAVR	15.71	42.94	84.29	97.16	99.50
PRO	11.78	75.79	85.96	93.60	94.64
PRIOR	14.34	50.63	84.29	97.41	99.67
BAYES	10.79	79.18	88.33	98.07	99.75
SVM POST	11.21	73.74	87.45	97.17	99.75
SVM BPOST	14.14	51.53	63.99	98.08	99.67
SVM BAYES	13.35	67.94	83.06	93.57	99.67

Table 5. Evaluation results for the mouth closed analysis our own face database

	EER	Detection Rate @ FAR			
		0.05	0.1	0.3	0.5
CL1	12.02	87.98	87.98	87.98	87.98
CL2	11.86	53.48	85.49	94.97	95.78
CL3	8.62	89.36	93.82	98.71	99.14
CL4	12.39	55.56	81.93	98.02	99.57
CL5	25.50	25.17	51.52	77.21	84.55
MAJ	3.61	97.00	97.00	97.00	97.00
MIN	16.03	83.76	83.86	84.24	84.62
MAX	13.04	40.26	47.47	100.00	100.00
AVR	3.61	97.85	97.85	99.14	100.00
BAVR	3.61	97.00	97.00	98.71	98.71
PRO	8.81	91.00	91.24	92.24	93.23
PRIOR	3.00	97.00	98.71	99.57	99.57
BAYES	2.26	97.85	98.82	99.57	99.57
SVM POST	3.00	98.28	98.71	99.57	99.57
SVM BPOST	3.00	97.00	98.71	99.57	99.57
SVM BAYES	3.11	97.67	98.78	99.57	99.57

tions, the most prominent one being aviation security. We specifically restrict our investigations to the *eyes-open* and *mouth-closed* criteria. Both criteria are challenging due to different possible input images showing noise, bad lighting, occlusions (glasses, beard) and the large variety of human faces in general. Our face analysis system builds upon several pre-processing steps before *eyes-open* and *mouth-closed* events are evaluated by different, complementary classification methods. We adopt a classifier fusion framework testing a number of different fusion strategies to combine the votes of different classifiers. In our experimental results we show that the classification performance on our criteria improves significantly due to classifier fusion, thus validating our hypothesis. The fusion strategies based on the learned linear combination utilizing the SVM approach show superior results, however, simpler fusion strategies exhibit competitive performance.

Based on our findings, further work is necessary to evaluate additional, complementary classification schemes to further improve the overall classification results for both criteria.

Acknowledgement

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

A. The Boosting Method

AdaBoost [19] is a well studied technique for supervised learning tasks. It has recently shown to be tremendously successful in a variety of object localization and classification applications. The most prominent work uses the boosted ensemble of a pool of weak classifiers for face detection [24]. The goal is to form a linear-combination of weak classifiers h_{weak} to get a strong-classifier $h_{strong} = \sum_i \alpha_i \cdot h_{weak,i}$, where α_i are the weights of each weak-classifier calculated during the training process. Weak classifiers are simple linear classifiers, which perform better than guessing, i.e., they are able to separate the training set with a classification rate larger than 50%. Examples for weak classifiers are Haar-like features [24] or edge orientation histograms [13]. Regarding the large number of potential weak classifiers that are tested in the training stage, efficient calculation becomes an important issue. This is solved using integral image representations on both Haar-like features and approximated edge orientation histograms [6]. Further aspects of techniques based on boosting weak classifiers are the use of cascade structures to speed-up localization tasks (localization can be seen as classification on a set of sliding windows moved over an image) and the type of boosting algorithm that is used to update the weights of weak classifiers during training, e.g., discrete, real or gentle AdaBoost.

B. The Active Appearance Model Method

The Active Appearance Model (AAM) [4] describes the variation in shape and texture of a training set representing an object. From a mean shape and a mean texture (which is defined in the coordinate frame of the mean shape) we calculate the modes of variation by applying PCA to shape, texture and to the combination of shape and texture, respectively. By keeping solely a certain percentage of the eigenvalue energy spectrum the model is represented very compactly and optimally regarding Gaussian noise. The AAM model fitting makes use of a learned regression model that describes the relationship between parameter updates and texture residual images according to [4]. Optimization takes place in a gradient descent scheme using the L2 norm of the intensity differences (between the estimated model and the given test image) as its cost function and the learned regression model for efficiently approximating the Jacobian of the cost function. A local minimum of the cost function corresponds to a model fitting solution. Since the minimum is local and the parameter space is very large, multi-resolution techniques have to be incorporated and the fitting requires coarse initialization.

C. The EigenFace Method

The *EigenFace* model [22] is a generative method that explicitly models face images in the face image space. It is based on registered image patches of equal dimensions and is calculated by applying a PCA to the face image vectors in the face image space. From this compact PCA representation, the eigenvectors according to the smallest eigenvalues are discarded. Given such a model consisting of a mean image and the remaining eigenvectors, it is possible to represent unknown image patches in the face image space. Thus, the reconstruction error of an image patch gives information about the similarity to facial images. By applying a threshold on the reconstruction error, classification can be performed.

References

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

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

[3] J. Canny. A computational approach to edge detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 8(6):679–698, 1986.

[4] 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.

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

[6] M. Grabner, H. Grabner, and H. Bischof. Fast approximated sift. In *Proc Asian Conf on Computer Vision (ACCV)*, pages 918–927, 2006.

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

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

[9] J. Kittler, M. Hatef, R. P. Duin, and J. Matas. On combining classifiers. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 20(3):226–239, March 1998.

[10] L. Kuncheva. A theoretical study on six classifier fusion strategies. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(2):281–286, 2002.

[11] S. Z. Li and A. K. Jain. *Handbook of Face Recognition*. Springer-Verlag New York, 2005.

[12] C.-J. Lin, R. C. Weng, and S. S. Keerthi. Trust region newton methods for large-scale logistic regression. In *Proceedings 24th International Conference on Machine Learning*, pages 561–568. ACM, 2007.

[13] D. G. Lowe. Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision*, 60(2):91–110, 2004.

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

[15] A. Martinez and R. Benavente. The AR face database. Technical Report 24, CVC, June 1998.

[16] N. Otsu. A threshold selection method from gray-level histogram. *IEEE Transactions on System Man Cybernetics*, SMC-9(1):62–66, 1979.

[17] M. Pantic and L. 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.

[18] P. Phillips, H. Wechsler, J. Huang, and P. Rauss. The FERET database and evaluation procedure for face recognition algorithms. *Image and Vision Computing*, 16(5):295–306, 1998.

[19] R. E. Schapire, Y. Freund, P. Bartlett, and W. S. Lee. Boosting the margin: A new explanation for the effectiveness of voting methods. *Ann. Stat.*, 26(5):16511686, 1998.

[20] 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.

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

[22] M. Turk and A. Pentland. Eigenfaces for recognition. *Journal of Cognitive Neuroscience*, 3(1):71–86, 1991.

[23] V. N. Vapnik. *The nature of statistical learning theory*. Springer-Verlag New York, 1995.

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

[25] J. Weickert, B. M. ter Haar Romeny, and M. A. Viergever. Efficient and Reliable Schemes for Nonlinear Diffusion Filtering. *IEEE Transactions on Image Processing*, 7(3):398–410, 1998.

[26] B. Wu and R. Nevatia. Improving part based object detection by unsupervised, online boosting. In *Proceedings IEEE Conference on Computer Vision and Pattern Recognition*, pages 918–927, June 2007.

[27] M. H. Yang, D. Kriegman, and N. Ahuja. Detecting faces in images: A survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(1):34–58, January 2002.

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