Fixed FAR Vote Fusion of Regional Facial Classifiers

L.J. Spreeuwers, R.N.J. Veldhuis, S. Sultanali, J. Diephuis

Faculty of Electrical Engineering, Mathematics and Computer Science
University of Twente, P.O. Box 217, 7500 AE Enschede, The Netherlands
l.j.spreeuwers@utwente.nl

Abstract: Holistic face recognition methods like PCA and LDA have the disadvantage that they are very sensitive to expression, hair and illumination variations. This is one of the main reasons they are no longer competitive in the major benchmarks like FRGC and FRVT. In this paper we present an LDA based approach that combines many overlapping regional classifiers (experts) using what we call a Fixed FAR Voting Fusion (FFVF) strategy. The combination by voting of regional classifiers means that if there are sufficient regional classifiers unaffected by the expression, illumination or hair variations, the fused classifier will still correctly recognise the face. The FFVF approach has two interesting properties: it allows robust fusion of dependent classifiers and it only requires a single parameter to be tuned to obtain weights for fusion of different classifiers. We show the potential of the FFVF of regional classifiers using the standard benchmarks experiments 1 and 4 on FRGCv2 data. The multi-region FFVF classifier has a FRR of 4% at FAR=0.1% for controlled and 38% for uncontrolled data compared to 7% and 56% for the best single region classifier.

1 Introduction

Automated face recognition basically started with the PCA (Principle Component Analysis) based eigenfaces method by Sirovich and Kirby [SK87] and Turk and Pentland [TP91]. The eigenfaces are the eigenvectors of the covariance matrix of the probability distribution of the vector space of face images. They form a basis set of facial images which are represented by linear combinations of eigenfaces. Facial images can be compared by various distance measures in the eigenface space. Linear Discriminant Analysis (LDA) based face recognition was introduced by Belhumeur et al. [BHK97]. LDA aims to find the combination of features that best separates between classes by maximising the ratio of the within class and the between class probabilities. To compare a probe sample $\mathbf{x}$ to a reference sample $\mathbf{y}$, this likelihood ratio becomes:

$$\text{LR}(\mathbf{x}, \mathbf{y}) = \frac{p(\mathbf{x}, \mathbf{y}|\text{same subject})}{p(\mathbf{x}, \mathbf{y}|\text{different subject})} \tag{1}$$

The likelihood ratio can be regarded as a score and the decision if two facial images are of the same subject is taken by comparing this score to a threshold. PCA and LDA are holistic methods. A deficiency of these holistic methods is that local variations in the face,

caused by facial expressions, occlusions by hair etc., see Figure 1, cannot be modelled well using normal distributions and there are generally insufficient training data available for those specific variations. The result is that performance decreases rapidly under facial expression, hair occlusions, illumination variations etc.

Figure 1: Local variations in the face due to caps, hair, illumination, glasses

One way to reduce sensitivity to these variations is to develop classifiers for parts of the face, called *experts*. Classifiers that combine multiple experts are called *multi-expert classifiers*. Examples of multi-expert face recognition can be found in [FYL+07] and [HAA07]. The combination or fusion of the experts is a multi-dimensional optimisation problem, which can be very complicated if the results of many experts must be fused. Popular fusion strategies are sum fusion, where the scores of the individual experts are simply added together to form a new score, and voting, where each expert may cast a vote and the number of votes forms a new score.

The multi-expert approach presented in this paper is based on LDA classifiers that operate on overlapping regions of the face which are combined by voting. If the appearance of the face is changed by expression or occlusion, only part of the face is affected while other parts remain stable, so some experts will still give a reliable result. Each expert casts a vote and hence a threshold must be defined for each expert. Our approach to finding the thresholds for the experts is based on choosing a fixed FAR for all experts. The Fusion of all experts we call, therefore, the *Fixed FAR Vote Fusion* or FFVF. We illustrated the basic ideas of this approach already for 3D face recognition in [Spr11]. In this paper we report the results of the FFVF approach for 2D face recognition.

The remainder of this paper is organised as follows: in section 2, the regional classifiers are described in detail. In section 3 the Fixed FAR Vote Fusion is explained. Section 4 briefly describes alignment of the images. Experiments and results are reported in section 5 and in section 6 conclusions are summarised.

2 Regional Classifiers

The LDA based classifier can be derived from the likelihood ratio given in equation 1, see [VB05]. We assume that the within distribution of all subjects is normal with the same within class covariance $\mathbf{C}_w$, but different means and that the total distribution of all faces is normally distributed with total covariance $\mathbf{C}_t$. A simple expression for the log of the likelihood ratio can now be calculated by first applying a transformation $\mathbf{T}$ that de-correlates and scales the total distribution such that it becomes white and simultaneously de-correlates the within distribution. This transformation is obtained using the singular

values and vectors of $\mathbf{C}_t$ and $\mathbf{C}_w$. The log of the likelihood ratio then becomes:

$$\log \mathrm{LR}(\mathbf{x}, \mathbf{y}) \propto (\mathbf{X} - \mathbf{Y})^T \Sigma_w^{-1} (\mathbf{X} - \mathbf{Y}) + \mathbf{X}^T \mathbf{X} + \mathbf{Y}^T \mathbf{Y} \tag{2}$$

Where $\mathbf{X} = \mathbf{Tx}$ and $\mathbf{Y} = \mathbf{Ty}$ the transformed sample vectors and Σ_w is a diagonal matrix, the within class covariance of the transformed face data. Some constants are ignored, because they merely offset and scale the log likelihood ratio score. Expression 2 deviates slightly from the commonly used expression for the likelihood ratio where the assumption is made that the reference sample $\mathbf{y}$ is the mean of the class. In this case the $\mathbf{Y}^T \mathbf{Y}$ term vanishes. Expression 2 gives the likelihood of the two samples being of the same class vs being of different classes, see [VB05]. Determining transformation $\mathbf{T}$ involves principle component analysis using Singular Value Decomposition (SVD) of the total distribution on a training set. Also a dimensionality reduction is applied and only the p largest singular values are kept. A second SVD is applied to the within class data to obtain Σ_w. The l smallest singular values are kept, that give the best discrimination between subjects.

As mentioned in the introduction, we combine many overlapping region classifiers using a voting strategy. If part of the face is affected by occlusion, expression, poor illumination conditions, the classifiers (experts) for those regions will give low scores. However, the classifiers for the unaffected or less affected regions will still give correct scores. Often, in order to obtain independent classifiers, non-overlapping regions are chosen for the experts. We chose larger, overlapping facial regions, resulting in dependent classifiers but with better performance. The regions were chosen such that each of them is stable for a certain kind of variation, e.g. there are regions excluding glasses, others excluding hair and yet others excluding the mouth area, which is very sensitive to expressions, or shaded eyes. We defined 30 regions which are illustrated in Figure 2.
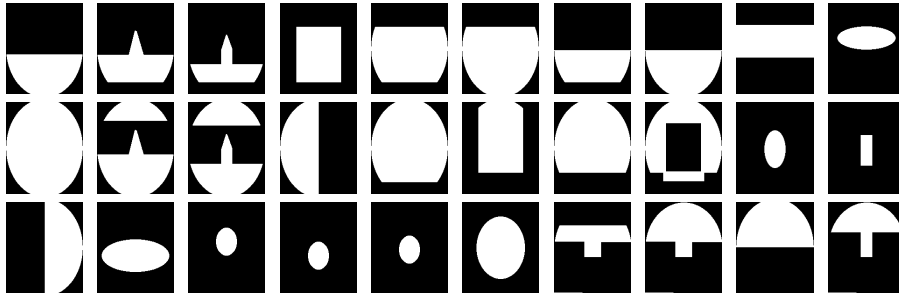

Figure 2: Regions used for different classifiers (experts).

3 Fixed FAR Vote Fusion

There are many approaches to fusion of multiple classifiers, but most of them, e.g. the sum rule, won't work well for dependent classifiers. Also, if e.g. one of the experts gives a very low score the resulting sum would be low even if the other experts give reasonably

high scores. A voting approach is much more robust in these cases. If two experts are dependent, they always give the same votes and therefore, the total score (sum of the votes) is simply offset by 1, which can easily be corrected by raising its threshold by 1. If one of the experts gives a very low score, the total vote count is still only decreased by a single vote. Therefore, we chose a voting scheme for fusion of the regional classifiers.

The next problem we have to solve is when each expert is allowed to cast a vote. Each expert E_i (region classifier) produces a log likelihood ratio score LR_i that is compared to an individual threshold T_i. If the score is above the threshold, the expert casts a vote V_i.

$$V_i = \left\{ \begin{array}{ll} 0, & \mathrm{LR}_i < T_i \\ 1, & \mathrm{LR}_i \geq T_i \end{array} \right. \tag{3}$$

The total number of votes S forms a new score that is the score of the fused experts. In its turn this score is compared to a threshold T to decide if the comparison gained sufficient support (accept) or not (reject).

$$D = \left\{ \begin{array}{ll} \text{reject}, & S = \sum_i V_i < T \\ \text{accept}, & S = \sum_i V_i \geq T \end{array} \right. \tag{4}$$

To find optimal thresholds T_i for all experts, is a complicated high dimensional optimisation problem. Rather than attempting to solve this problem exhaustively, we simplify the problem. We don't want poor experts to contribute too much to the errors of the fused classifier. If a poor classifier often gives a high score for a comparison of two images of different subjects, the threshold for this expert should be high. Likewise, for good classifiers, we choose lower thresholds. This desired behaviour can be realised by choosing the thresholds T_i such that all experts have the same False Accept Rate (FAR). The fixed FAR value for each expert is now a single parameter with which we can tune the performance of the fused classifier. We call this approach Fixed FAR Vote Fusion (FFVF).

4   Registration

Accurate registration or face alignment is of great importance for proper comparison of facial images using LDA. We used a landmark based approach with 4 landmarks: left and right eyes, nose and mouth. The landmarks were detected automatically using Viola & Jones' boosted cascade detectors [VJ01] within the rectangle of the face area, which was also detected using a boosted cascade detector. For each face with detected landmarks, a transformation consisting of rotation, scaling and translation was derived to map the landmarks to a set of mean landmark positions that was obtained using Procrustes analysis on a set of 500 faces. Registration worked well for both images acquired under controlled as well as uncontrolled circumstances. In addition to registration, we also performed histogram equalisation or Anisotropic Smoothing [GB03] to the images. The former worked better for images recorded in controlled circumstances, while the latter gave better results for uncontrolled recordings.

5 Experiments and Results

5.1 Data and experimental setup

The performance of the FFVF multi-expert classifier was evaluated using the FRGC (Face Recognition Grand Challenge) benchmarks for 2D face recognition [PFS$^+$06]. This database contains $50\,000$ recordings of a total of 466 subjects. Each session consists of 4 controlled images, two uncontrolled images and 1 3D image. The controlled images were recorded in a studio setting and were full frontal facial images with two different illuminations and facial expressions. The uncontrolled images were taken at various locations with different illumination and facial expressions. A total of 6 experiments were defined for these data. In this paper, we refer to experiments 1 and 4 which address 2D face recognition for controlled and uncontrolled circumstances, see Figure 3.
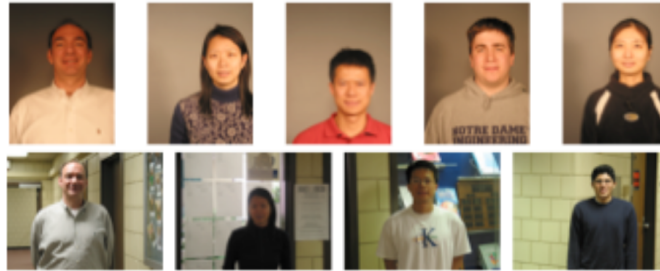

Figure 3: Top row: controlled facial recordings from FRGC; bottom row: uncontrolled

In experiment 1, single controlled 2D images are used as reference and probe samples. A total of $12\,769$ images of 222 subjects is available for training consisting of controlled and uncontrolled images, but we only used controlled images for training. The evaluation set for experiment 1 consists of $16\,028$ controlled images of 466 subjects. In experiment 4, a single controlled image is used as reference and compared to a single uncontrolled image. The training-set consists of the same $12\,769$ images as in experiment 1. Here we used both controlled and uncontrolled images for training. The evaluation set for experiment 4 consists of $16\,028$ controlled reference images and $8\,014$ uncontrolled probe images. All results are reported as verification rate (VR) at a false accept rate (FAR) of 0.1%.

After registration, the images were re-sampled to a size of 64×45 pixels. A lower resolution reduces execution times, but may also impact performance. The chosen size turned out to be a good compromise between the two.

Next the following experiments were carried out:

- Determination of the optimal number of PCA and LDA coefficients (p and l)
- Results for experiment 1 and 4 for single regions
- Determination of the optimal setting for the Fixed FAR
- Results for experiment 1 and 4 for the fused classifier

5.2 PCA and LDA coefficients

We investigated the optimal settings for the number of PCA and LDA coefficients by determining the FRR at FAR=0.1% for the training-set. For 3 different experts: mask 0, 9 and 29, the FRR as a function of the number of PCA and LDA coefficients are given in Figure 4. The optimal setting for the number of PCA coefficients is above 160 and for LDA above 45. We chose 200 and 65 in our further experiments.
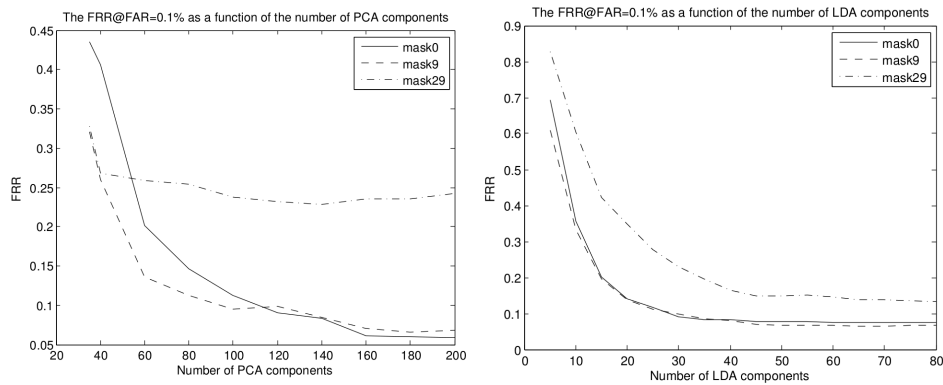

Figure 4: FRR@FAR=0.1% for varying numbers of PCA and LDA coefficients.

5.3 Results for single regions

Next, using the two evaluation sets for experiments 1 and 4 (controlled and uncontrolled), the FRR at FAR=0.1% was determined for the individual classifiers (experts). The FRR for experiment 1 for the different experts varied from 7% to 58% and for experiment 4 from 56% to 95%. In Table 1, the results are given for the full face mask and the best performing mask experts. For completeness, also the Equal Error Rates (EER) are provided.

	full face mask		best mask	
	EER	FRR@FAR=0.1%	EER	FRR@FAR=0.1%
Experiment 1	2.1%	8.3%	1.9%	7.2%
Experiment 4	8.5%	56.2%	8.5%	56.2%

Table 1: Performance of individual experts

The best performing mask for the controlled images is the full face with the mouth area removed. For the uncontrolled images (experiment 4) it is just the full face mask.

5.4   Determination of the Fixed FAR

The Fixed FAR to determine the thresholds for the experts is the single parameter of the FFVF fusion. To find its optimal value, the training set was split into 2 partitions, one for training and one for tuning the Fixed FAR (7666 resp. 5110 images of 222 subjects). All experts were retrained using the training partition and the threshold for each expert was determined for a range of Fixed FAR values. Next the performance of the fused classifier was determined by finding the threshold on the number of votes for FAR=0.1% on the tuning partition. The performance of the fused classifier on the tuning partition expressed as FRR at FAR=0.1% as a function of the Fixed FAR is given in Figure 5.
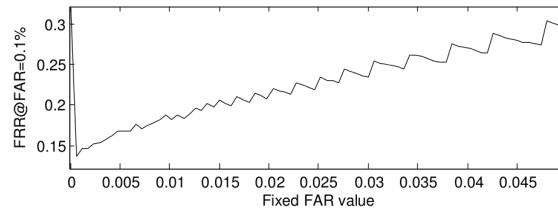

Figure 5: Performance of the fused classifier as a function of the Fixed FAR

From Figure 5 we can observe that the optimal choice for the Fixed FAR is 0.1% (lowest FRR).

5.5   Results for fused classifier

Finally, the experts are retrained on the full training-set and the thresholds are determined on the training-set for a Fixed FAR of 0.1% as found in section 5.4 and fused classification results are determined for the controlled and uncontrolled evaluation-sets as defined for experiments 1 and 4 of the FRGC benchmarks. The results, expressed as FRR at FAR=0.1% and EER are given in Table 2 together with the best performing individual experts.

	fused classifier		best single expert	
	EER	FRR@FAR=0.1%	EER	FRR@FAR=0.1%
Experiment 1	1.4%	4.0%	1.9%	7.2%
Experiment 4	9.2%	38.4%	8.5%	56.2%

Table 2: Performance of fused classifier

We observe a significant reduction of the FRR at FAR=0.1% for both controlled (7.2% to 4.0%) as well as uncontrolled images (56.2% to 38.4%). The EER for the fused classifier for experiment 1 (controlled) decreases, but for experiment 4 (uncontrolled) slightly increases relative to the best single expert. The latter can be explained by the fact that the fused classifier was optimised for FAR=0.1%.

6   Conclusion

We present a multi-expert approach to face comparison with many classifiers for over-lapping regions. The reasoning is that if part of the face is affected due to variations in expression, illumination or occlusion, other parts remain stable and can be used for reliable facial recognition. We propose a Fixed FAR Vote Fusion (FFVF) approach, where each expert casts a vote if its score is above a threshold which is selected by the Fixed FAR. The number of votes is regarded as the score of the fused classifier. We show that using regional LDA based classifiers fused using FFVF, performance improves for the controlled (FRR drops from 7.2% to 4% at FAR=0.1%) and uncontrolled (FRR drops from 56.2% to 38% at FAR=0.1%) experiments of the FRGC 2D face recognition benchmark.

References

[BHK97]   P. N. Belhumeur, J. P. Hespanha, and D. J. Kriegman. Eigenfaces vs. Fisherfaces: Recognition Using Class Specific Linear Projection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19(7):711–720, July 1997.

[FYL+07]   Yun Fu, Junsong Yuan, Zhu Li, T.S. Huang, and Ying Wu. Query-Driven Locally Adaptive Fisher Faces and Expert-Model for Face Recognition. In *IEEE International Conference on Image Processing ICIP2007*, volume 1, pages I – 141–I – 144, Sept 2007.

[GB03]   R. Gross and V. Brajovic. An Image Preprocessing Algorithm for Illumination Invariant Face Recognition. In *Proceedings of the 4th International Conference on Audio- and Video-based Biometric Person Authentication*, AVBPA'03, pages 10–18, Berlin, Heidelberg, 2003. Springer-Verlag.

[HAA07]   M.T. Harandi, M.N. Ahmadabadi, and B.N. Araabi. A hybrid model for face recognition using facial components. In *Signal Processing and Its Applications, 2007. ISSPA 2007. 9th International Symposium on*, pages 1–4, Feb 2007.

[PFS+06]   P. J. Phillips, P. J. Flynn, T. Scruggs, K. W. Bowyer, and W. Worek. Preliminary face recognition grand challenge results. In *Proceedings of the 7th International Conference on Automatic Face and Gesture Recognition FGR 2006*, pages 15–24, 2006.

[SK87]   L. Sirovich and M. Kirby. Low-dimensional procedure for the characterization of human faces. *Journal of the Opical Society of Ameraca A*, 4:519–524, 1987.

[Spr11]   L.J. Spreeuwers. Fast and Accurate 3D Face Recognition Using Registration to an Intrinsic Coordinate System and Fusion of Multiple Region classifiers. *International Journal of Computer Vision*, 93(3):389–414, March 2011. Open Access.

[TP91]   M. Turk and A. Pentland. Eigenfaces for Recognition. *Journal of Cognitive Neuroscience*, 3(1):71–86, January 1991.

[VB05]   R. N. J. Veldhuis and A. M. Bazen. One-to-template and ono-to-one verification in the single- and multi-user case. In *26th Symposium on Information Theory in the Benelux, Brussels, Belgium*, pages 39–46, Brussels, May 2005.

[VJ01]   P. Viola and M. Jones. Rapid object detection using a boosted cascade of simple features. In *Computer Vision and Pattern Recognition, 2001. CVPR 2001. Proceedings of the 2001 IEEE Computer Society Conference on*, volume 1, pages I–511–I–518 vol.1, 2001.