Automatic Face Alignment by Maximizing Similarity Score

Bas Boom, Luuk Spreeuwers and Raymond Veldhuis

University of Twente, EEMSC, Signals & Systems
P.O. box 217, 7500 AE, Enschede, The Netherlands

Abstract. Accurate face registration is of vital importance to the performance of a face recognition algorithm. We propose a face registration method which searches for the optimal alignment by maximizing the score of a face recognition algorithm. In this paper we investigate the practical usability of our face registration method. Experiments show that our registration method achieves better results in face verification than the landmark based registration method. We even obtain face verification results which are similar to results obtained using landmark based registration with manually located eyes, nose and mouth as landmarks. The performance of the method is tested on the FRGCv1 database using images taken under both controlled and uncontrolled conditions.

1 Introduction

Several papers have shown that correct registration is essential for good face recognition performance [1],[2]. The performance of popular face recognition algorithms, for instance PCA, LDA and ICA, depend on accurate face registration. We propose a new method for face registration which searches for the optimal face alignment by maximizing the score of a face recognition algorithm. Our new method outperforms the landmarks methods described in [3]. In this paper we investigate the practical usability of the new face registration method for face verification.
In practice, we need to locate the face region using a face detection algorithm. Using this region, we register the face image to a user template in the database and then recognize the face. Our face registration algorithm, first described in [4], finds an optimal face alignment from the located face region. In this paper we investigate different practical aspects of our face registration method. We test our face registration method with different face classifiers as evaluation criteria in the search procedure. We test our method under circumstances where lighting is controlled and uncontrolled, and we also lower the resolution of the face images. We investigate if our method works with automatically registered training images, so it becomes fully automatic. Finally, we look at the mistakes of our registration method and introduce some solutions to overcome these problems.
In the literature, face registration is usually achieved by finding landmarks in a face image. An approach which is similar to our face registration algorithm is described in [5] and [6], which uses a form of robust correlation to find the alignment to a user template. Recently, Wang et al [7] improve the face identification on the the FERET database by

Boom B., Spreeuwers L. and Veldhuis R. (2007).
Automatic Face Alignment by Maximizing Similarity Score.
In *Proceedings of the 7th International Workshop on Pattern Recognition in Information Systems*, pages 221-230
DOI: 10.5220/0002427602210230

calculating the similarity score of different alignments and selecting the maximal score. The main differences with their approach are the assumption of a face identification problem and using the manually labelled eye coordinates as start points.
In section 2 we firstly explain our method. Secondly, we specify our search procedure and finally we describe the face recognition algorithms. In section 3, we describe the experimental setup we used to evaluate our method. Section 4 describes the various experiments carried out using our registration method. The final section gives a conclusion about this face registration method.

2 Matching Score based Face Registration

We have developed a new face registration method, namely Matching Score based Face Registration (MSFR). This method searches for the optimal alignment between the probe image and a user template in the database. To evaluate the alignment, we use the output of a face classifier. This output is also called the matching score in the case of a genuine user or the similarity score in the case of a unknown user. We assume that the similarity score becomes higher if the alignment of face image to the template image improves. The second assumption is that the optimal alignment of the genuine user's face image gives a higher similarity score than the optimal alignment of an imposter's face image.

2.1 Face Registration

Based on the assumptions described above, we have developed the following method. The region of the face is found by a face detection algorithm, in our case the face detector first described by Viola and Jones [8]. Using an affine transformation $T_{\boldsymbol{\theta}}$ on the pixel p of the probe image I_p, given by the region found by the face detection algorithm, we vary the multiple registration parameters $\boldsymbol{\theta}$ searching for the optimal alignment. The geometric transformation function is :

$$\begin{aligned} T_{\boldsymbol{\theta}}(x, y) = (&\theta_1 cos\theta_2 x - \theta_1 sin\theta_2 y + \theta_3, \\ &\theta_1 cos\theta_2 x + \theta_1 sin\theta_2 y + \theta_4) \end{aligned} \tag{1}$$

In the transformed image $I_p(T_{\boldsymbol{\theta}}(p))$, the pixel values are calculated using bilinear interpolation. The optimal alignment parameters to person i in the database are given by:

$$\boldsymbol{\theta}_{max} = \arg\max_{\boldsymbol{\theta} \in \Theta} S(I_p(T_{\boldsymbol{\theta}}(p)), i) \tag{2}$$

Of course, the similarity score $S(I_p(T_{\boldsymbol{\theta}_{max}}(p)), i)$ can also be used to verify the person's identity. It is also possible to use one face recognition algorithm to find the optimal alignment parameters but another face recognition algorithm to classify the face.

2.2 Search for Maximum Alignment

To search for the maximum similarity score, we use a search algorithm called the downhill simplex method [9]. This search algorithm finds four parameters $\boldsymbol{\theta}$ which maximize the similarity score. The starting point of the search algorithm is the region given by the face detection algorithm, where $\boldsymbol{\theta}_0 = (1, 0°, 0, 0)$. For the downhill simplex method, we need to determine a simplex (geometrical figure in N dimensions consisting of $N+1$ points). This is created from the starting point parameters and four points for which we varied a single parameter. The other four points of the simplex are: $\boldsymbol{\theta}_1 = (1.2, 0°, 0, 0)$, $\boldsymbol{\theta}_2 = (1, 5°, 0, 0)$, $\boldsymbol{\theta}_3 = (1, 0°, 5, 0)$ and $\boldsymbol{\theta}_4 = (1, 0°, 0, 5)$. We have also experimented with other simplexes to start the search algorithm. Details will be given later on in this paper. We have also experimented with the search algorithm of Powell-Brent [10],[11], but it performs worse for this search problem.

2.3 Face Recognition Algorithms

Face recognition involves performing several steps to be able to recognize a face in an image. Using an aligned image $I_p(T_{\boldsymbol{\theta}}(p))$ given by the search algorithm, we select a region of interest (ROI) and we normalize the image inside the ROI to zero mean, unit variance. After that, the pixels in the ROI are vectorized and the resulting vectors are then used in our face recognition algorithm.
We use four algorithms to calculate the similarity score, these algorithms are based on PCA [12] some in combination with LDA [13]. In this paper, we used a fixed number of PCA and LDA dimensions, 100 and 50 respectively. The first algorithm is PCA in combination with the Euclidean distance (eucl), where we calculate the Euclidean distance between the probe image with the template in the database and use this as similarity score. The second algorithm is PCA in combination with the Mahalanobis distance (mah), where we use the Mahalanobis distance instead of the Euclidean distance. In the third algorithm, we perform feature reduction using PCA and LDA and use the log-likelihood ratio proposed in [14] to calculate the similarity score. For a certain class i, the similarity score S is calculated by:

$$\begin{aligned} S_{\boldsymbol{y},i} = &-(\boldsymbol{y} - \boldsymbol{\mu}_{W,i})^T \Sigma_W^{-1} (\boldsymbol{y} - \boldsymbol{\mu}_{W,i}) \\ &+\boldsymbol{y}^T \Sigma_T^{-1} \boldsymbol{y} - \log|\Sigma_W| + \log|\Sigma_T| \end{aligned} \tag{3}$$

Where $\boldsymbol{y}$ is a vector which is a representation of the face image after feature reduction, Σ_T is the total covariance matrix, Σ_W is the within class covariance matrix and $\boldsymbol{\mu}_{W,i}$ is the ith class average. The final algorithm uses the numerator of the likelihood ratio, which is given by:

$$S_{\boldsymbol{y},i} = -(\boldsymbol{y} - \boldsymbol{\mu}_{W,i})^T \Sigma_W^{-1} (\boldsymbol{y} - \boldsymbol{\mu}_{W,i}) - \log|\Sigma_W|$$

The reason behind using only the numerator is that we register to a certain user template. This means that it is not necessary to maximize with respect to the background distribution, given by the denominator in the likelihood ratio. We call this final method the within ratio. This method is only intended for finding a maximal alignment to a user template and not for face recognition. After the face registration, a final face verification is always performed using the likelihood ratio.

3 Experimental Setup

In our experiments, we use the Face Recognition Grand Challenge (FRGC) version 1 database [15]. We only used face images in which the face was correctly found by the face detection algorithm of Viola-Jones [8], because we are not interested in the mistakes made during face detection. The face images are correctly found when the eyes, nose and mouth coordinates lie inside the face region and the width and height of this region are less then four times the distance between the eyes. The FRGC version 1 database contains 275 individuals, from which we use a set of 3761 face images taken under controlled conditions and a set of 1811 face images taken under uncontrolled conditions. In our experiments we randomly split these sets into two subsets, each consisting of approximately half of the images of each person. One subset is used for training and enrollment and the other is used for testing. The same random split is used for all experiments.
We compare our face registration method with the best landmark based face registration methods in [3], namely MLLL + BILBO. The results of the face registration are measured on the performance in face verification by calculating Equal Error Rate (EER): this is the point of operation where the False Accept Rate (FAR) is equal to the False Reject Rate (FRR). To measure the accuracy of registration we use the RMS error. We calculated the location of the eyes, nose and mouth in the original image based on the alignment found by MSFR. The RMS error is then calculated between these positions and the manually labelled landmarks given by the FRGC database and we normalize to a distance of 100 pixels between the eyes.

4 Experiments

Since our earlier paper [4], we have performed more extensive experiments. We have done several experiments to gain a better understanding of our method. In our first experiment we report the results of the different recognition algorithms on the FRGC database. In the second experiment, we use a lower resolution making the algorithm faster and applicable to video surveillance environments. The third experiment investigated if this face registration algorithm can be trained on face images which are registered using automatically obtained landmarks. The final experiments try to address some failures of the search algorithm by performing the search several times and adding registration noise to the training data.

4.1 Comparison between Recognition Algorithms

Searching for the best alignment requires a recognition algorithm. In this section, we describe our experiments using the various recognition algorithms. We compare the results with both manually labelled landmarks and automatically obtained landmarks. For our experiment, we use the experimental setup described in section 3. We use face images with a resolution of 128×128 pixels. Training the face classifier is achieved using a training set which is aligned using the manually labelled landmarks. Face recognition applied to images registered using MLLL + BILBO, however, is train also on images which were registered using MLLL + BILBO [3]. After registering the face, we recognize the registered faces using the Likelihood ratio classifier.

Table 1. Results of the face verification using various registration algorithms.

	FRGC Controlled			FRGC Uncontrolled		
Registration Method	EER [%]	RMS error users	RMS error impostors	EER [%]	RMS error users	RMS error impostors
Manually labelled	0.59	-	-	1.7	-	-
MLLL+BILBO	3.6	7.9	7.9	9.7	10.2	10.2
PCA eucl	1.8	3.4	9.5	3.2	4.3	10.3
PCA mah	1.3	3.0	5.7	**1.5**	2.9	4.2
Likelihood ratio	**1.01**	3.2	9.4	2.3	4.0	7.9
Within ratio	**1.07**	3.2	8.7	2.1	3.8	7.2

In Table 1, we compare the results of the various registrations in EER and RMS error. The columns for the EER show that the MSFR outperforms the landmark registration. By comparing the various classifiers of MSFR, it becomes clear that the likelihood ratio and the within ratio perform best on the controlled images of FRGCv1, although the performance of PCA with Mahalanobis distance is also very good. On the uncontrolled image of FRGCv1, PCA with Mahalanobis distance performs best, even better than the manually labelled landmarks. In RMS error, the various MSFR methods are more accurate than MLLL+BILBO when it comes to registration to the genuine user. If we look at registration of an imposter, the RMS error of most of the MSFR is higher than the RMS error of MLLL + BILBO. This does not need to have any effect on the EER, because poorly registered images usually do not improve the similarity score. In the case of PCA with Mahalanobis distance, the RMS error of the impostor is still lower than that of MLLL+BILBO.

Figure 1 shows the FAR and FRR curves of the manually labelled landmarks and the MSFR approaches on the controlled images of the FRGC. Both the matching and non-matching scores increase when using MSFR, which means that for genuine users, better alignments than manually labelled landmarks can be found using MSFR.

4.2 Lowering Resolution

In [4] we report that our method takes about 20-30 seconds to register and classify a face image using an Intel Pentium 2.80 GHz. Currently, it takes about 5-10 seconds on the same computer for a face image of 128×128 pixels, because we optimized some of our source code. In [16] we investigated the effect of the image resolution on face recognition. It turns out that the EER does not increase much on face recognition by

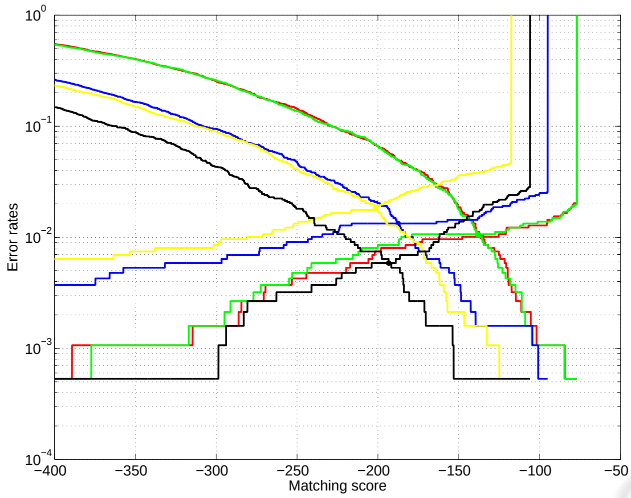

Fig. 1. FAR and FRR curves: the red line is the likelihood ratio, the green line is within ratio, the blue line is PCA mah, the yellow line is PCA eucl and the black line is the manual registration.

lowering the resolution to 32×32. In practice, we do not always have high resolutions face images, so we performed a experiment at a resolution of 32×32 pixels, which also leads to a decrease in computation time. The results of the recognition are stated in Table 2 together with the results of using the normal resolution of 128×128 pixels.

Table 2. The EERs by using face images with a resolution of 32×32.

Registration Methods	FRGC Controlled resolution 128×128	FRGC Controlled resolution 32×32	FRGC Uncontrolled resolution 128×128	FRGC Uncontrolled resolution 32×32
PCA eucl	1.8	2.4	3.2	5.5
PCA mah	1.3	2.3	1.4	2.8
Likelihood ratio	1.01	1.3	2.3	3.8
Within ratio	1.07	1.7	2.1	3.6

Although Table 2 shows that the EER increases somewhat by using face images of 32×32 pixels, these results are still acceptable and better than face registration using MLLL + BILBO. It takes about 2-5 seconds to register and classify a face image of 32×32 pixels on a Intel Pentium 2.80 GHz. More improvements in the operation time of our method can be realised, because we have not payed much attention to this subject yet.

4.3 Training using Automatically Obtained Landmarks

Until now, we have assumed that for training and enrollment of the face registration we can use a set of manually labelled face images. In practice, this usually is not the case, especially for the enrollment of a new user. This is the reason we performed an experiment where we trained and enrolled images which have been aligned using the landmarks given by MLLL + BILBO. The results of this experiment are given in Table 3.

Table 3. The EERs when the training en enrollment are registered using MLLL + BILBO on images from the controlled set of FRGCv1.

Registration Methods	EER [%] manual	EER [%] automatic
PCA eucl	1.8	2.0
PCA mah	1.3	1.5
Likelihood ratio	1.01	1.7
Within ratio	1.07	1.8

Although the results we report in table 3 show increased error rates, the performance is still much better than using MLLL + BILBO for face registration. This shows that correct registration of training set is not critical.

4.4 Improving Maximization

In this paper we already reported a large improvement in the results of face registration. But after some analysis of our method, we found that a correct registration is not always found by simply running the downhill simplex search algorithm. The main reason is that the search algorithm can find a local maximum far away from the global maximum. In figure 2, we show the incorrectly found registration results by the likelihood ratio classifier. These results can easily be determined by considering the RMS error of face images. The faces depicted in figure 2 all have a RMS error bigger than 11 pixels, except for the bottom right face image. The main reason for these errors is that in these cases, the search algorithm searches in the wrong direction and gets stuck in a local maximum.

To correct these outliers, we have developed two strategies to address this problem. Firstly, we use the downhill simplex method several times but start with a different simplex in the search space. Secondly, we change the search space by training on a database with some registration noise.

Using a Different Start Simplex The first strategy is based on the idea that if we start searching from another side in the search space, we will probably never come across the same local maximum. For this experiment, we have defined two new start simplexes; the first simplex consists of the points: $\boldsymbol{\theta}_0 = (1, 0°, 0, 0)$, $\boldsymbol{\theta}_1 = (0.8, 0°, 0, 0)$, $\boldsymbol{\theta}_2 = (1, -5°, 0, 0)$, $\boldsymbol{\theta}_3 = (1, 0°, -5, 0)$ and $\boldsymbol{\theta}_4 = (1, 0°, 0, -5)$, so that we start from the opposite side of the search space. For the second simplex, we start at the points: $\boldsymbol{\theta}_0 = (0.9, -2.5°, -2.5, -2.5)$, $\boldsymbol{\theta}_1 = (1.1, -2.5°, -2.5, -2.5)$, $\boldsymbol{\theta}_1 = (0.9, 2.5°, -2.5, -2.5)$,

Fig. 2. Face images which have been incorrectly registered, the bottom-right image is an example of a correct alignment.

$\boldsymbol{\theta}_2 = (0.9, 2.5°, -2.5, -2.5)$, $\boldsymbol{\theta}_3 = (0.9, -2.5°, -2.5, 2.5)$ and $\boldsymbol{\theta}_4 = (0.9, -2.5°, -2.5, 2.5)$, which gives us a search area around the located face region.

Table 4. The EERs when searching from different positions in the search space on images from the controlled set of FRGCv1.

Registration Methods	start position 1	start position 2	start position 3	combining 1,2,3
PCA eucl	1.8	7.4	3.5	3.6
PCA mah	1.3	1.6	5.2	**0.64**
Likelihood ratio	1.01	2.6	6.7	**0.59**
Within ratio	1.07	2.4	6.5	**0.64**

In Table 4 we present the results of the three different start positions. The EERs 2 and 3 in table 4 are the new start positions, while EER 1 gives the start position which has been used throughout the entire paper. Also, the results of combining these outcomes of registration by using the maximum similarity score of the three different start positions is given in table 4, this procedure is done for both genuine users and impostors. These combinations give results which are similar to registration using manually labelled landmarks. The EERs of the other start positions are not particularly good, but using other starting points results in different failures. By using the similarity score to evaluate the final outcomes of the different starting points, the local maxima are discarded.

Adding Noise to Train Our Registration Method The second strategy is based on changing the search space. This is done by adding gaussian noise to the manually labelled landmarks of the training examples. By adding noise to the training, we also hope that we can model the registration error better. The results of this experiment are given in Table 5 where the t is the standard deviation of the noise in pixels, normalized to 100 pixels between the eyes.

Table 5. The EERs when adding noise to the registration of the training on controlled face image of FRGCv1.

Registration Methods	noise						combining $t = 0, 1, 2$
t	0	1	2	3	4	5	
PCA eucl	1.8	1.5	1.5	1.8	1.3	1.5	1.9
PCA mah	1.3	0.91	0.85	0.80	0.69	1.2	**0.80**
Likelihood ratio	1.01	0.69	0.75	0.91	0.91	1.3	**0.59**
Within ratio	1.07	0.64	0.75	0.91	0.96	1.2	**0.64**

In Table 5 we show that by combining the results of adding registration noise to the training set we reach the same result as with manually labelled landmarks. Another observation is that by adding a little registration noise to the training, the EER seems to decrease anyway. This is because a reduction in the number of outliers. We suspect that the registration noise makes the search space smoother in the areas further away from the optimal registration. By adding too much noise, the EER increases. This can be observed for $t = 5$ in table 5.

5 Conclusion

In this paper, we present a system for face registration which uses the output of the face recognition classifier to find an optimal registration. We search for an optimal registration by varying the face alignment parameters. Our new face registration method performs better than the landmark based methods of [3]. The experiments show that our method performs well with both face images taken under controlled as well as uncontrolled conditions. The operating speed of the method has been improved and we show that lowering the resolution improves the speed even more while still obtaining good performance. Our face registration method also works with an automatically registered training set and achieves good results despite registration errors in the training set. This makes our face registration method very useful in practise when dealing with a face verification problem. By using multiple searches, the results of our face registration method are equivalent to the results obtained with registration using manually labelled landmarks. This kind of performance has not yet been achieved by any fully automatic face registration method known to us.

References

1. Riopka, T., Boult, T.: The eyes have it. In: Proceedings of ACM SIGMM Multimedia Biometrics Methods and Applications Workshop., Berkeley, CA (2003) 9–16
2. Beumer, G., Bazen, A., Veldhuis, R.: On the accuracy of EERs in face recognition and the importance of reliable registration. In: SPS 2005, IEEE Benelux/DSP Valley (2005)
3. Beumer, G., Tao, Q., A.M.Bazen, Veldhuis, R.: A landmark paper in face recognition. 7th International Conference on Automatic Face and Gesture Recognition (FGR06) (2006) 73–78
4. Boom, B., Beumer, G., Spreeuwers, L., Veldhuis, R.: Matching score based face registration. In: PRORISK 2006, STW (2006)
5. Jonsson, K., Matas, J., Kittler, J., Haberl, S.: Saliency-based robust correlation for real-time face registration and verification. In: Proc British Machine Vision Conference BMVC98. (1998) 44–53
6. Matas, J., Jonsson, K., Kittler, J.: Fast face localisation and verification. In: Image and Vision Computing. Volume 17. (1999) 575–581
7. Wang, P., Tran, L.C., Ji, Q.: Improving face recognition by online image alignment. In: 18th International Conference on Pattern Recognition (ICPR 2006). Volume 1. (2006) 311–314
8. Viola, P.A., Jones, M.J.: Rapid object detection using a boosted cascade of simple features. In: CVPR (1). (2001) 511–518
9. Nelder, J., Mead, R.: A simplex method for function minimization. The Computer Journal **7** (1965) 308–313
10. Brent, R.: Algorithms for Minimization without Derivatives. Prentice-Hall, Englewood Cliffs, N.J. (1973)
11. Powell, M.: An efficient method for finding the minimum of a function of several variables without calculating derivatives. Computer Journal **7** (1964) 155–162
12. Turk, M., Pentland, A.: Eigenfaces for recognition. Journal of cognative neuroscience (1991) 71–86
13. Belhumeur, P.N., Hespanha, J., Kriegman, D.J.: Eigenfaces vs. fisherfaces: Recognition using class specific linear projection. In: ECCV 2. (1996)
14. Veldhuis, R., Bazen, A., Booij, W., Hendrikse, A.: Hand-geometry recognition based on contour parameters. In: Proceedings of SPIE Biometric Technology for Human Identification II, Orlando, FL, USA (2005) 344–353
15. NIST: Frgc face db. (http://www.frvt.org/FRGC/)
16. Boom, B., Beumer, G., Spreeuwers, L., Veldhuis, R.: The effect of image resolution on the performance of a face recognition system. In: ICARCV'06. (2006)