Comparing landmarking methods for face recognition

G.M. Beumer, Q. Tao, A.M. Bazen and R.N.J. Veldhuis
g.m.beumer@utwente.nl
University of Twente, EEMSC, Signals & Systems
P.O. box 217, 7500 AE, Enschede, The Netherlands

Abstract—**Good registration (alignment to a reference) is essential for accurate face recognition. We use the locations of facial features (eyes, nose, mouth, etc) as landmarks for registration. Two landmarking methods are explored and compared: (1) the Most Likely-Landmark Locator (MLLL), based on maximizing the likelihood ratio [1], and (2) Viola-Jones detection [2]. Further, a landmark-correction method based on projection into a subspace is introduced.**

Both landmarking methods have been trained on the landmarked images in the BioID database [3]. The MLLL has been trained for locating 17 landmarks and the Viola-Jones method for 5 landmarks. The localization error and effects on the equal-error rate (EER) have been measured. In these experiments ground- truth data has been used as a reference. The results are described as follows:

1. The localization errors obtained on the FRGC database are 4.2, 8.6 and 4.6 pixels for the Viola-Jones, the MLLL, and the MLLL after landmark correction, respectively. The inter-eye distance of the reference face is 100 pixels. The MLLL with landmark correction scores best in the verification experiment.
2. Using more landmarks decreases the average localization error and the EER.

Keywords: facial feature, face registration, face recognition, landmarking, likelihood ratio, Viola-Jones, landmark correction

I. INTRODUCTION

Research by Riopka et al. [4], Cristinacce et al. [5] and Beumer et al. [6] showed that precise landmarks are essential for a good recognition performance. Cristinacce did research on landmark locators based on correlation, orientation maps and Viola-Jones based algorithm [7]. The Viola-Jones algorithm was proposed by Viola et al. [2] In this paper we propose an improvement on earlier work by Bazen et al. [1] and a Viola-Jones based landmark finder. Both methods will be compared to each other and the groundtruth data. The methods will be compared by calculating the RMS value of the error between the results and to the groundtruth data. Also the outcome of an verification experiment measured by the equal error rate (EER) will be given.

II. LANDMARK DETECTION

The first step in face recognition is to locate the face in the image. This is done with a Viola-Jones based algorithm [2] from the OpenCV library [8]. We assumed that there is only one face per image. When the face is found a region of interest (ROI) is selected for each landmark. In this ROI we search for landmarks using one of the two algorithms. Both will be explained in this section.

A. Most Likely Landmark Location

In MLLL landmark finding is seen as a two-class classification problem: a location in an image is either the landmark or it is not. The texture values in a region surrounding a landmark are used as features for the classification. For each location in the ROI the likelihood ratio -for that location to be the landmark- is calculated. The *most* likely location, i.e. the one with the highest score, is assumed to be the landmark. The landmarks used can be seen in Figure 1. Outliers may occur due to errors by the MLLL. Sometimes these outliers can be corrected using a shape correction. Both the landmark detection and the shape correction are discussed in more detail in the next sections.

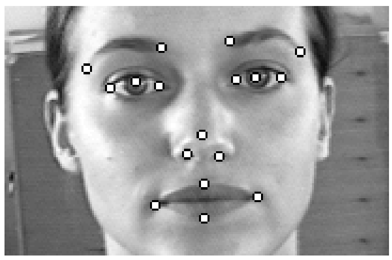

Fig. 1. Landmarks in the face used

A.1 Likelihood ratio based facial feature finder

The MLLL calculates the likelihood ratio for a landmark at each position in the ROI. The gray level intensities of the neighbourhood of a candidate location (u,v) form a vector $x_{u,v}$. The likelihood that $x_{u,v}$ is the neighbourhood of the landmark is found by calculating

$$L(u,v) = \frac{p(x_{u,v}|L)}{p(x_{u,v}|\bar{L})} \tag{1}$$

The location of the landmark is chosen at the point (u, v) at which $L(u, v)$ is maximum. This is done for all 17 landmarks.

A.2 Shape correction

Sometimes landmarks are detected incorrectly. The aim of shape correction is to detect or correct the errors made. This can be done by defining a space in which correct shapes fit but deformed ones will not. A shape is the collection of the coordinates of a set of landmarks. Correct shapes are assumed to be in a subspace of $\mathbb{R}^{2d}$ with d the number of landmarks. A basis of the subspace is determined by means of principal component analysis (PCA). The shape is projected *there and back again* (BILBO). In the subspace the number of features is reduced. Projecting it back to the original space coordinates which not fit the model will have shifted. The landmarks of which the location changed significantly during BILBO, are considered to be wrong. The new location is taken as the correct location.

To train BILBO we need a set of shapes. The shapes consist of 17 coordinates of landmarks in the face. A shape vector consists of a column vector with all the u coordinates and all the v coordinates making x, 34×1 in size. A model is trained on groundtruth data which are put as columns in a matrix X.

Training BILBO consists of the following steps:
1. Register all the shapes in X, to the average shape.
2. Add -limited- rotation, translation and scaling to all shapes in X in order to model variance encountered in the images. We added normal distributed variations. The translation has a standard deviation of 5 pixels. The scale has a standard deviation of 5%. Finally the rotation has a standard deviation of 3 degrees.
3. Perform an SVD on X: $X = USV^T$.
4. Reduce features by taking only the first n columns of U.
5. Calculate a transformation matrix; $T = UU^T$.

To correct a shape the next algorithm is used:
1. Estimate the shape after transformation, $x' = Tx$.
2. Determine the Euclidian distance D_i per landmark between x and x'.
3. Determine threshold τ. The threshold is calculated as

$$\tau = RC\frac{1}{17}\sum_{i=1}^{17} D_i \tag{2}$$

where C is a constant and R is the run number.
4. Only for coordinates of which $D_i > \tau$: update vector x: $x_i = x'_i$.
5. Repeat steps 1 to 4. Once for a landmark $D_i < \tau$ stop updating it until no $D_i > \tau$ for all i.
6. Repeat step 1 to 5 changing all coordinates until $R = 5$.
7. Transform the coordinates back to the original scale.

In Figure 2 an example of corrected landmarks is shown. The circles show the raw landmarks and the triangles show the new location of the landmarks.

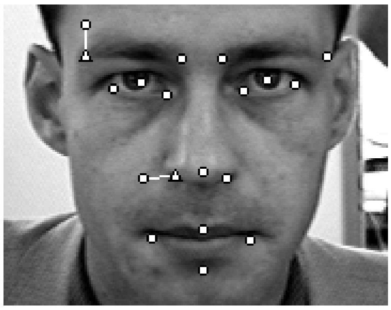

Fig. 2. Original shape and corrected landmarks (triangles)

B. *Viola-Jones based facial landmark localization*

The second approach for landmark localization is the Viola-Jones based method [2], which uses a combination of Haar-like features to represent the texture information in an image. A detailed description of this method and the training method -Adaboost- can be found in [2].

We developed 5 landmark detectors, namely two eyes (size 28×14), one nose (size 28×14), and two mouth corners (size 20×20). The locations of the templates (positive samples) are obtained from the BioID ground truth data, as shown in Figure 1. The negative samples are randomly chosen from the background which does not contain the landmarks. Compared to the MLLL method, only 5 landmarks are chosen for Viola-Jones based method, because the texture of other landmarks (like eyebrows and lips) are not as constant as these 5 landmarks for AdaBoost training. Experiments also showed that they could not result in fast and compact cascades for detection.

The BioID database contains 1,521 images which have been manually landmarked. From each image we obtaine 5 facial landmark templates.The code of the Adaboost training is taken from the Intel OpenCV library [2]. In our work each detector is trained with 3,000 positive samples and 6,000 negative samples. For simplicity the face region is firstly detected as ROI to localize landmarks in face. Figure 3 shows the results of applying the Viola-Jones method for localizing face and landmarks. This method does a multi-scale search and chooses the facial landmark candidates through thresholding [2]. There is the possibility of multiple candidates (multi-size and/or different position), or missing candidate for one facial landmark. For the multi-size case, we choose the candidate with the largest size. The reason is that it is observed that smaller-size

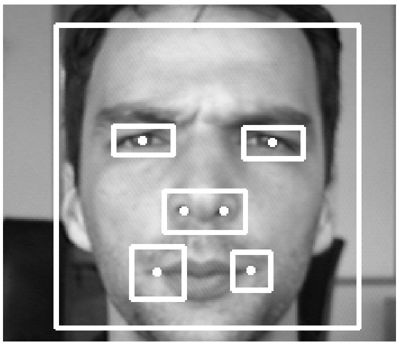

Fig. 3. The landmarking result by Viola-Jones method

candidates tend to be less reliable, like a small eye candidate localized on the eyebrow. For candidates with the same size but in difference positions, we calculate their relative locations in reference to the face ROI, and choose the one nearest to its average position. In case of missing landmarks, we use fewer points for registration. Besides, it is also possible to restore the missing landmarks according to the statistics of their geometrical distribution in face region.

III. Experiments and results

Two methods for the localization of landmarks were proposed. Both will be evaluated in two ways. First we evaluate the RMS distance between the found landmark location and the groundtruth. Then the landmarks are used for registration in a verification experiment. The EER will serve as a benchmark for the quality of the landmarks.

For the experiments we used two databases. The BioID [3] is used to train the feature detection and the shape correction. The BioID database consists of 1521 images which vary in pose, scale and lightning conditions, but are mainly frontal. The FRGC database [9] was used to test the alignment algorithms. The FRGC database was also used to train and test the verification experiment. The FRGC database consists 5658 images. Two third, 3772, of the database are high quallity images with low variety in pose, lighting and scale with around 300 pixels inter eye distance. For our tests we used only the high quality part of FRGC database images.

A. Training the MLLL

The MLLL detector is trained on the 17 landmarks as shown in Figure 1. The positive templates were all selected using hand labeled groundtruth data. The negative samples were all taken around the landmark at a minimal distance of half the size of the template. All templates are either 40x40 pixels or 60x40 pixels in size. The templates are all normalized in energy. This means that they are zero mean and have standard deviation one. For training the feature extraction we used the method proposed by Bazen et al. [10], which uses AMDA for feature reduction. The feature extraction matrix T is stored as are the covariance matrices and averages.

B. Accuracy of the facial feature detection

To evaluate the landmarking we need a well defined measure for error. Since there are images of various scales in the image a simple root mean square (RMS) distance is not sufficient. In order to calculate a meaning full measure a simple method was used:
1. Translate, scale and rotate the groundtruth data so that the eyes ate at 100 pixels apart.
2. Align the shape found to the groundtruth shape.
3. Calculate the distance between each landmark and it's groundtruth equivalent.
4. Remove bias which is caused by the differences in labeling policy between the databases i.e. tip of the nose versus a point between the nostrils in the different databases.
5. Calculate the RMS value of the remaining difference between the found shape and the groundtruth shape.

This was done on the FRGC database. The labels from the FRGC database the center of the mouth while our methods label the mouth corners. In this case the mouth corners were averaged to make an estimate of the center of the mouth.

B.1 Results

The results can be found in Table I. This table shows that BILBO works. The RMS error on the landmark location improves a lot. After BILBO the MLLL performs approximately the same as the Viola-Jones method. On the FRGC data again both perform equally well on the eyes. On the nose and the mouth the BILBO+MLLL outperforms the Viola-Jones method. The results on the nose suggest that the Viola-Jones method needs more training.

C. Impact on verification experiment

In this section, the verification experiment we used for face recognition is discussed. The images are aligned us-

FRGC	Right eye	Left eye	Nose	Mouth
Viola-Jones	3.2	3.3	6.3	4.1
MLLL	6.7	7.2	13.0	7.3
BILBO+MLLL	4.2	4.6	5.8	3.7

TABLE I
RELATIVE RMS ERROR RESULTS ON THE FRGC DATABASE.

ing the found coordinates. Half the images are used to train the verification while the other half is used to test it. The experiments are repeated 10 times in order to get a good estimate of the expected EER of the data for this algorithm. More information on the algorithm can be found in in an article by Beumer et al. [6]. To investigate the impact of the number of landmarks on the registration we also did experiments on only 5 out of 17 landmarks for the MLLL coordinates. The eyes, mouth corners and the nose were used for this.

It should be noted that the verification experiment is not optimized and that parameter tuning might improve its overall performance. However, this is not necessary in order to compare results. The experiment uses the coordinates found by one of the two algorithms or the groundtruth data to register the faces.

C.1 Results

The results can be seen in Table II. It shown that using groundtruth verification gives good results (EER = 0.45%). Our methods results in higher EERs around 4%. The results for the MLLL and BILBO+MLLL show that the shape correction does improve the results. Also they both outperform the Viola-Jones method. The experiment on only 5 landmarks shows however that this difference is caused by the fact that only 5 landmarks are available for the Viola-Jones method and 17 for the MLLL.

IV. Conclusions

The proposed methods for registration are not yet accurate enough to compete with groundtruth data. The proposed shape correction work but only when enough landmarks are available. The Viola-Jones method performs best in terms of absolute distance. For the verification experiment it is lacking. This is due to the fact that it only is trained on 5 landmarks. It ourperforms the MLLL when only using 5 landmarks. The MLLL with shape correction performs best. It is however to be expected that when the number of landmarks for the Viola-Jones method can be extended this method will outperform the MLLL also on the verification experiment. This is also due to the fact that a shape correction method then can be used. The results from experiments by Beumer et al. [6] which show that more landmarks are good for noise reduction are confirmed by this research.

FRGC	EER[%]	std(EER)[%]
Ground truth data	0.45	0.03
Viola-Jones	4.9	0.1
MLLL	4.0	0.1
BILBO+MLLL 17 landmarks	3.6	0.1
BILBO+MLLL 5 landmarks	6.1	0.1

TABLE II
RESULTS OF THE BIOMETRIC VERIFICATION EXPERIMENT

References

[1] A.M. Bazen, R.N.J. Veldhuis, and G.H. Croonen, "Likelihood ratio-based detection of facial features," in *Proc. ProRISC 2003, 14th Annual Workshop on Circuits, Systems and Signal Processing*, Veldhoven, The Netherlands, nov 2003, pp. 323–329.

[2] Paul Viola and Michael Jones, "Robust real-time object detection," *International Journal of Computer Vision*, 2002.

[3] HumanScan, "Bioid face db," `http://www.humanscan.de/`.

[4] T.P. Riopka and T. Boult, "The eyes have it," in *Proceedings of ACM SIGMM Multimedia Biometrics Methods and Applications Workshop.*, Berkeley, CA, 2003, pp. 9–16.

[5] D. Cristinacce, T. Cootes, and I. Scott, "A multi-stage approach to facial feature detection," in 15^{th} *British Machine Vision Conference, London, England*, 2004, pp. 277–286.

[6] G.M. Beumer, A.M.Bazen, and R.N.J. Veldhuis, "On the accuracy of eers in face recognition and the importance of reliable registration.," in *SPS 2005*. IEEE Benelux/DSP Valley, April 2005.

[7] D. Cristinacce and T. Cootes, "A comparison of shape constrained facial feature detectors," in 6^{th} *International Conference on Automatic Face and Gesture Recognition 2004, Seoul, Korea*, 2004, pp. 375–380.

[8] Intel, "Open computer vision library," `http://sourceforge.net/projects/opencvlibrary/`.

[9] NIST, "Frgc face db," `http://www.frvt.org/FRGC/`.

[10] A.M. Bazen and R.N.J. Veldhuis, "Detection of cores in fingerprints with improved dimension reduction," in *Proc. SPS 2004*, Hilvarenbeek, The Netherlands, apr 2004, pp. 41–44.