Automatic Landmark Detection and Face Recognition for Side-View Face Images

Pinar Santemiz
Luuk J. Spreeuwers
and Raymond N.J. Veldhuis
Signals and Systems Group, Department of Electrical Engineering
University of Twente
Drienerlolaan 5 P.O.Box 217
7500AE Enschede, The Netherlands
p.santemiz@utwente.nl
L.J.Spreeuwers@utwente.nl
R.N.J.Veldhuis@utwente.nl

***Abstract*—In real-life scenarios where pose variation is up to side-view positions, face recognition becomes a challenging task. In this paper we propose an automatic side-view face recognition system designed for home-safety applications. Our goal is to recognize people as they pass through doors in order to determine their location in the house. Here, we introduce a recognition method, where we detect facial landmarks automatically for registration and identify faces. We test our system on side-view face images from CMU-Multi PIE database. We achieve 95.95% accuracy on detecting landmarks, and 89.04% accuracy on identification.**

I. Introduction

In applications dealing with identifying people from videos such as surveillance systems or smart homes, face recognition is the primary biometrics. One possible application area for face recognition are home-safety applications. Here, face recognition can be used to increase the situational awareness, and to prevent the factors that may cause further accidents. However, in real-life scenarios with uncontrolled environment, face recognition becomes a challenging task due to occlusion, expression, or pose variations.

In this paper we introduce a novel method for side-view face recognition to be used in house safety applications. Our aim is to identify people as they walk through doors, and estimate their location in the house. We design a system that uses video recordings from cameras attached to door posts under ambient illumination. The cameras have a limited view angle thus preserving the privacy of the people. Here, we test our system in a setting similar to this scenario. We use multiple still images that contain side-view face images, and we perform automatic landmark detection and recognition tests on these images.

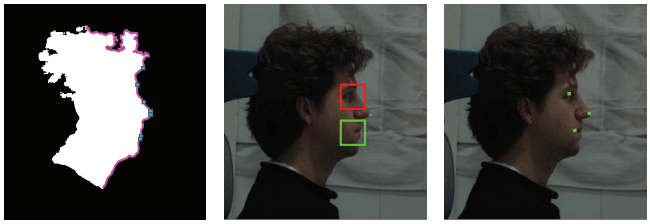
Due to the complex structure of human face, face recognition under pose variation up to side-view is a difficult problem. In [1], a literature survey on face recognition under pose variations can be found. In initial attempts to compare side-view face images, mainly profile curves or fiducial points on the profile curves were used. One such method is proposed by Bhanu and Zhou [2], where they find nasion and throat point, and compare the curvature values using Dynamic Time Warping (DTW). They achieve a recognition accuracy of 90% on Bern database, which contains side-view face silhouettes of 30 people.

In video-based applications, people make use of the texture information in addition to profile curves. Tsalakanidou *et al.* [3] present a face recognition technique where they use the depth map for exploiting the 3D information, and apply Eigenfaces. They experiment on the XM2VTS database using 40 subjects, and recognize 87.5% of them correctly. In a recent study [4], Santemiz *et al.* proposes a side-view face recognition method using manual landmarks. Here, they use local binary patterns to compare faces and achieve a recognition accuracy of 91.10% on a small subset from CMU-Multi PIE database [5], where they excluded the subjects wearing glasses.

In this study, we first find three landmark points automatically using Histogram of Oriented Gradients (HOG) [6] and train Support Vector Machines (SVM) [7]. We use these landmarks for registering images as presented in Section II. Then we apply Principal Component Analysis (PCA) [8], Linear Discriminant Analysis (LDA) [9], Local Binary Pattern (LBP) [10], and Histogram of Oriented Gradients (HOG) [6] to describe the face images. The details of our feature extraction techniques are given in Section III. We identify faces using nearest neighbor classifier and test our system on side-view face images of CMU-Multi PIE database [5]. We analyze our results in Section IV. Finally, we will give our conclusion in Section V, and discuss our future work.

II. Automatic Landmark Detection and Registration

In our landmark detection approach, we aim to find three landmark points on the face, namely, the eye center, the tip of the nose, and the corner of the mouth. A visualization of these landmark points is given in Figure 1(c). In our method we first manually select the skin color masks from training samples containing 50 subjects and 708 images, and learn the multivariate Gaussian distribution of the HSV color space. Using this distribution, we estimate the skin color masks of the remaining images, and extract the outer profile. Then, we compute the curvatures on facial profile, and use the curvatures



(a) (b) (c)

Fig. 1. Landmark detection steps. (a) Found skin color mask, the outer profile, and the candidate points for the nose tip. (b) ROIs for the eye center and mouth corner for one nose tip candidate. (c) Automatically detected landmarks.

(a) (b) (c) (d)

Fig. 2. Automatic landmark detection results. The green points are ground truth, and the red points are the found landmark locations.(a) -90 degrees, (b) -75 degrees, (c) 75 degrees, (d) 90 degrees

having a local maxima as candidate points for the tip of the nose which is shown in Figure 1(a). Around each candidate point we extract a Region of Interest (ROI) of size 55×55 pixels. We assume that the ROI of the eye and the mouth is centered at a distance of $[-40, +40]$ pixels and $[-40, -40]$ pixels away from the tip of the nose, respectively. An example is given in Figure 1(b). For each nose tip candidate, we extract ROIs for the mouth corner and eye center, and scan all three ROIs to find the landmarks. An example is shown in Figure 1.

To find the landmark points, we train three separate SVMs. In training, for each landmark location we select nine positive and 16 negative samples of image patches of size 10×10 pixels. To select the positive samples, we use the manually labeled coordinates and eight neighboring coordinates for each landmark. The negative samples are chosen randomly from the ROIs of the landmarks. From all these image patches we extract the HOG features and train SVMs. Here we use the same training set as we use for training the skin colors. Using the SVMs, we compute scores for each candidate point and choose the three coordinates having the total maximum score as our landmarks.

For registration, we use Procrustes analysis [11] to find the transformation parameters between each image. First, we align the landmarks of the images in the training set to the landmarks of the first image, and compute their mean to find the average landmarks. Then, we compute the transformation between each image and the average landmarks, and transform images, accordingly. Finally, in order to have fixed sized images we place a bounding rectangle around the face, and crop the image. Here, we use a fixed window for the bounding rectangle of size 200×100 pixels such that the right side of the rectangle is centered at the tip of the nose. Some examples can be seen in Figure 2.

III. Feature Extraction

We describe the registered face images using two baseline algorithms, Principal Component Analysis (PCA) and Linear Discriminant Analysis (LDA), and also using Local Binary Pattern (LBP), and Histogram of Oriented Gradients (HOG).

PCA [8] (Eigenface approach) is an algorithm for reducing dimensionality of a feature space by projecting it onto a space that spans the significant variations, and LDA [9] (Fisherface approach) is a supervised method for classification problems. In our implementation, we learn the PCA parameters from the training set, project each image into PCA space, and from the projected values of the training samples we learn LDA parameters. For classification, we use nearest neighbor method using cosine similarity measure.

Local Binary Pattern is a method that describes the local spatial structure of an image [10]. The most prominent advantages of LBP are its invariance against illumination changes, and its computational simplicity. In our system, we divide the images into 75 subregions, and compute the LBP histograms for each region. Then, we concatenate these histograms to form the feature vector of the image. For classification, we use nearest neighbor method using Chi square distance measure.

Histogram of Oriented Gradients (HOG) are mainly used in computer vision as feature descriptors in object detection and recognition [6]. HOG represents the shape via the distributions of local intensity gradients or edge directions. The main advantage of using HOG descriptors is that they offer some robustness to scene illumination changes, while capturing characteristic edge or gradient structure. We divide the image into cells with 10×10 pixels and for each cell, we form an orientation histogram having 32 bins. For classification, we use nearest neighbor method using Chi square distance measure.

IV. Experimental Results

In CMU Multi-PIE database, each subject is recorded under 15 poses in up to four sessions, where 13 cameras are located at head height spaced at 15 degrees intervals. The images are acquired in a controlled environment with constant background and illumination, and have a resolution of 640×480 pixels. We select the images acquired from the four cameras that are located at $-90, -75, 75$,and 90 degrees as side-view images and use a total of 3684 side-view face images from all 337 subjects in our experiments.

A. Landmark Detection

In our landmark detection experiments, we divide the set indo two subsets: a training set containing 50 subjects and 708 images, and a test set with 287 subject and 2976 images. The average distance between the eye center and the mouth corner in this set is 79.91 pixels. Therefore, an automatically detected point displaced 10-pixels distance from the ground truth is accepted as a correct detection. Using this threshold we detect 95.95% of the landmarks correctly, where the correct detection for the eye center, the tip of the nose, and the mouth corner separately are 94.79%, 96.34%, and 96.72%, respectively.

In our experiment, our skin color segmentation algorithm failed to detect the face in only one image where the subjects face is mostly covered by hair as seen in Figure 3(a). Other than this example, we were able to segment the skin color masks, but had cluttered profile curve on some images due to hair, facial hair, eyeglasses, or poor illumination. Yet, our approach to eliminate the false candidates using HOG and SVM proved to be successful in most of the examples.

When we observe the 363 images where our algorithm falsely detected landmarks, we see that the errors are mostly caused by occluded images due to hair or eyeglasses. Yet we also observe that in some images our algorithm falsely detect the upper lip location as the tip of the nose. Some false landmark detection examples are given in Figure 3.

(a) (b) (c) (d)

Fig. 3. False landmark detection examples. The green points are ground truth, and the red points are the found landmark locations. (a) Failed skin color segmentation. (b) Falsely detected nose tip. (c) Occlusion of hair. (d) Occlusion of eyeglasses.

B. Recognition

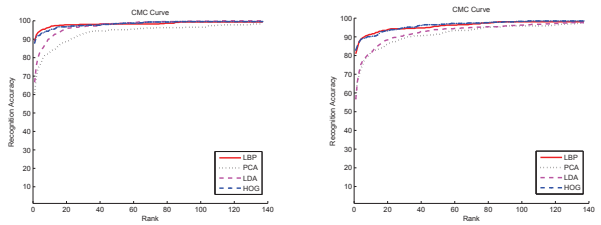
In our identification experiments, we divide the database into three subsets: a training set containing 200 subjects and 2484 images, an enrollment set with 137 subjects and a total of 744 images consisting of six images for each subject, and a test set with a total of 456 images. Since we aim to use side-view face recognition to identify people from video recordings, here we keep the setting much similar to this scenario, and use multiple still images for enrollment. The enrollment images and the test images can have a 15 degrees pose variation, which we expect to be the case in a real life scenario.

We perform identification experiments using PCA, LDA,LBP, and HOG. We further applied sum-rule fusion to LBP and HOG. We test our recognition method on images registered using only the tip of the nose, using three manually labeled landmarks, and using automatically detected landmarks. Our rank-one accuracies can be seen in Table I, and the Cumulative Match Characteristic (CMC) curves for identification in Figure 4.

TABLE I. RANK 1 IDENTIFICATION PERFORMANCES

	Registered using One Manual Landmark	Registered using Three Manual Landmarks	Registered using Three Automatic Landmarks
PCA	61.18%	60.96%	56.80%
LDA	62.06%	66.67%	56.58%
LBP	82.89%	88.82%	80.92%
HOG	85.75%	87.94%	82.89%
LBP+HOG	85.53%	89.04%	82.02%

When using one landmark we achieve our best performance using HOG features and obtain 85.75% recognition accuracy.



(a) (b)

Fig. 4. Cumulative Match Characteristic (CMC) curves. (a) CMC curve achieved on images registered using manually labeled landmarks, (b) CMC curve achieved on images registered using automatically detected landmarks.

Our highest accuracy for images registered with three manual landmarks is 89.04% which we obtain using sum-rule fusion of LBP and HOG. For images registered with automatically detected landmarks our best performance is 82.89% which is obtained using HOG features.

When we analyze these results, we see that LBP and HOG consistently perform better than PCA and LDA. It has been shown that compared to holistic methods, LBP is less sensitive against variations that occur due to illumination, expression, or pose [10]. Both HOG, and LBP describe the image by dividing it into local regions, extracting texture descriptors for each region independently, and then combining these descriptors to form a global description of the image. Consequently, they are not effected by small local changes as much as PCA or LDA. When we compare HOG and LBP, we see that they achieve similar results for each registration method. Also when we look at the CMC curves, we see that on higher ranks both LBP and HOG have similar results. However LBP is more effected by errors of automatic landmark detection which shows that HOG copes with local changes slightly better than LBP.

When we compare identification results of registered and not registered images we see that we achieve better results with registered images except PCA. The results improve much significantly for LBP compared to HOG, which supports the robustness of HOG against local changes compared to LBP.

We observe that our recognition accuracies drop significantly when we use automatically detected landmarks. To better understand the cause of this decline we perform another experiment using the samples for which the landmark is correctly found. For these samples images registered using manual landmarks give a rank-one recognition accuracy of 86.43%, where as using images registered with automatic landmarks the rank-one recognition accuracy increases to 87.62%. Based on this observation, we conclude that finding the landmarks within 10 pixels is accurate enough, and the decline we see in performances is caused by the samples whose landmarks are falsely detected.

In order to better analyze these results, we also investigate the erroneous cases. Some misclassification examples caused by false landmark detection, and occlusion of hair or glasses can be seen in Figure 5. We observe that the misclassification errors for LBP and HOG are very similar based on the type of errors. We show two misclassification errors of LBP in Figures 5(a) and 5(b), and two misclassification errors of HOG in Figures 5(c) and 5(d).

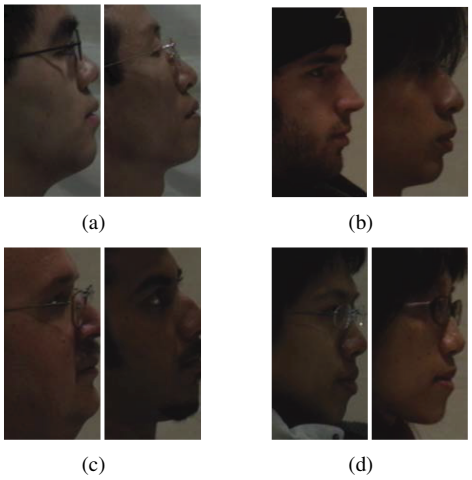


Fig. 5. Misclassification examples: the test images (left), the nearest images found by the classifier (right). (a) and (b) Misclassification examples of LBP. (c) and (d) Misclassification examples of HOG. (a) and (c) Misclassification due to falsely detected nose tip. (b) Misclassification due to hair. (d) Misclassification due to glasses.

In examples shown in Figures 5(a) and 5(c), the landmark detection algorithm falsely detects the upper lip location as the tip of the nose, and the faces are tilted upwards. We see that the pose and the shape of the faces are similar, but the difference in texture is significantly different. Especially, in the example shown in Figure 5(c), the test sample wears glasses and does not have beard, which is the opposite for the sample that is found as the most similar. When we compare the samples shown in Figure 5(b), the test sample wears a hat and the found sample has his forehead covered with hair in a similar way. In Figure 5(d), in both images the left eye of the sample is partly shown which shows that they both have the same head pose. Also, both the test sample and the found sample wear glasses.

V. Conclusion and Future Work

In this work we investigate automatic landmark detection and side-view face recognition to be used in house safety applications, where we aim to identify people as they walk through open doors, and estimate their location in a house. Here, we present our initial results that we achieved using side-view face images from the CMU-Multi PIE database. We automatically detect the landmarks with a detection accuracy of 95.95% and use these landmark points for registration. We test our system both with manually labeled landmarks and automatically detected landmarks using PCA, LDA, LBP, and HOG. We achieve 89.04% recognition accuracy using sum-rule fusion of LBP and HOG for manually labeled landmarks, and 82.89% recognition accuracy using HOG for automatically detected landmarks.

We see that, our automatic landmark detection method is effective, and shows high accuracy. Also, when we compare identification results using the samples for which our algorithm detects landmarks correctly, we see that the performance using automatic landmarks is higher than the performance using manual landmarks. Moreover, we achieve promising results with our recognition algorithm.

In the future, we aim to improve our landmark detection algorithm and increase the number of landmarks to better cope with images that are partially occluded due to hair or glasses. We also aim to include a higher pose variation in our experiments.

Acknowledgment

This work is supported by GUARANTEE (ITEA 2) 08018 project.

References

[1] X. Zhang and Y. Gao, "Face recognition across pose: A review," *Pattern Recognition*, vol. 42, no. 11, pp. 2876–2896, November 2009. [Online]. Available: http://dx.doi.org/10.1016/j.patcog.2009.04.017

[2] B. Bhanu and X. Zhou, "Face recognition from face profile using dynamic time warping," *Int. Conf. on Pattern Recognition (ICPR)*, vol. 4, pp. 499–502, 2004. [Online]. Available: http://dx.doi.org/10.1109/ICPR.2004.1333820

[3] F. Tsalakanidou, "Use of depth and colour eigenfaces for face recognition," *Pattern Recognition Letters*, vol. 24, no. 9-10, pp. 1427–1435, June 2003. [Online]. Available: http://dx.doi.org/10.1016/S0167-8655(02)00383-5

[4] P. Santemiz, L. Spreeuwers, and R. Veldhuis, "Side-view face recognition," in *WIC Symposium on Information Theory in the Benelux*, 2011, pp. 305–312.

[5] R. Gross, I. Matthews, J. Cohn, T. Kanade, and S. Baker, "Multi-PIE," *Image and Vision Computing*, vol. 28, no. 5, pp. 807–813, May 2010. [Online]. Available: http://dx.doi.org/10.1016/j.imavis.2009.08.002

[6] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in *Proc. IEEE Conf. CVPR*, vol. 2, 2005, pp. 886–893.

[7] V. V. Corinna Cortes, "Support-vector networks," *Machine Learning*, vol. 20, pp. 273–297, 1995.

[8] M. Turk and A. Pentland, "Eigenfaces for Recognition," *Journal of Cognitive Neuroscience*, vol. 3, no. 1, pp. 71–86, January 1991. [Online]. Available: http://dx.doi.org/10.1162/jocn.1991.3.1.71

[9] P. Belhumeur, J. Hespanha, and D. Kriegman, "Eigenfaces vs. Fisherfaces: recognition using class specific linear projection," *IEEE Trans. on PAMI*, vol. 19, no. 7, pp. 711–720, 1997. [Online]. Available: http://dx.doi.org/10.1109/34.598228

[10] T. Ahonen, A. Hadid, and M. Pietikainen, "Face Description with Local Binary Patterns: Application to Face Recognition," *IEEE Trans. on PAMI*, vol. 28, no. 12, pp. 2037–2041, 2006. [Online]. Available: http://dx.doi.org/10.1109/TPAMI.2006.244

[11] C. Goodall, "Procrustes methods in the statistical analysis of shape," *J. Royal Statistical Society, Series B (Methodological)*, p. 285339, 1991.