IET Biometrics

Research Article

IET Journals
The Institution of Engineering and Technology

Face reconstruction from image sequences for forensic face comparison

ISSN 2047-4938
Received on 10th June 2015
Revised on 28th September 2015
Accepted on 12th October 2015
doi: 10.1049/iet-bmt.2015.0036
www.ietdl.org

Chris van Dam ✉, *Raymond Veldhuis, Luuk Spreeuwers*

Faculty of EEMCS Chair of Services, University of Twente, Cybersecurity & Safety P.O. Box 217, 7500 AE Enschede, The Netherlands
✉ *E-mail: c.vandam@utwente.nl*

Abstract: The authors explore the possibilities of a dense model-free three-dimensional (3D) face reconstruction method, based on image sequences from a single camera, to improve the current state of forensic face comparison. They propose a new model-free 3D reconstruction method for faces, based on the Lambertian reflectance model to estimate the albedo and to refine the 3D shape of the face. This method avoids any form of bias towards face models and is therefore suitable in a forensic face comparison process. The proposed method can reconstruct frontal albedo images, from multiple non-frontal images. Also a dense 3D shape model of the face is reconstructed, which can be used to generate faces under pose. In the authors' experiments, the proposed method is able to improve the face recognition scores in more than 90% of the cases. Using the likelihood ratio framework, they show for the same experiment that for data initially unsuitable for forensic use, the reconstructions become meaningful in a forensic context in more than 60% of the cases.

1 Introduction and background

Face comparison in a forensic context is a challenging task. Reconstructing faces with a model driven approach [1–13] have been addressed by several researchers, but always suffers from a bias towards the underlying face model. For example, a face model trained on multiple look-a-like faces of a person is far more likely to match with that person than a general face model. In a forensic context avoiding any form of bias towards other faces is crucial to create better opportunities for evidence in court. In this paper, we explore the possibilities of a dense model-free three-dimensional (3D) face reconstruction method, based on an image sequence from a single camera. In the first stage, we use landmarks in multiple images to obtain a coarse estimation of the shape of a face and the rotation and translation parameters of the face in each frame. In the second stage we enhance the reconstruction by estimating the reflection coefficient and surface normals simultaneously for every point in the 3D shape. On the basis of the estimate of the normals, we enhance the 3D shape of the face. Applying these steps iteratively leads to a dense 3D reconstruction of a face including texture information based on the reflection coefficients of the face. The proposed method is model-free and has therefore no bias towards an underlying face model. We further refer to this as the model-free requirement. As a consequence of this requirement, the reconstruction process becomes more difficult, since no average model can be used as a starting point for the 3D reconstruction. Moreover, where, for example, the Morphable model method can provide reconstructions based on a single image [1], we need multiple images. Our goal is to increase the performance of face recognition, while still maintaining the model-free requirement of the forensic context.

Although some parts of our work are related to stereo and multi-camera reconstruction approaches, see, for example [14], our problem is more complex. First, in a stereo setup the exact positions and viewing directions of the cameras are known, while we need to estimate the position and direction of each view. Estimations are less precise and depend on the quality of the data. Second, there is deformation of the faces due to expression and motion artefacts. In a multi-camera setup this could never happen, because in such a case all images are recorded at the same moment in time. Third, the constant brightness assumption [15] is not applicable in a single camera situation, because the object is moving, and so the illumination chances in each frame. Therefore, there is more uncertainty during the reconstruction process. Garg *et al.* [16] introduce a variational non-rigid structure from motion approach, but they use a massive number of tracking points, which is not realistic using forensic data. In their experiments they use synthetic face sequences, where the level of difficulty for reconstruction is far less than with real image sequences. Delaunoy and Pollefeys [17] present a photometric bundle adjustment method for multi-view 3D modelling which shows low reprojection errors, but all their experiments are performed on feature rich objects including unrealistic views in comparison to faces under pose.

In our reconstruction process, we handle a moving face in front of a static camera, so the constant brightness assumption is not valid in our case. We introduce a reconstruction process based on image sequences which avoids any bias towards face models. Our reconstruction method provides dense 3D reconstructions of the face sequences.

The remaining of this paper is structured as follows: First our proposed method is introduced in Section 2, including a dense reconstruction algorithm. In Section 3, we included face comparison experiments on frontal reconstructions, the forensic implications of the proposed reconstruction method and visual inspection of the dense 3D reconstructions. Section 4 concludes our paper.

2 Method

The proposed method is designed to work with real image data, comparable to image sequences obtained from an ATM camera. Multiple non-frontal images under varying pose are available, captured from a moving face with a single camera. Due to this setup the illumination is different in every image. Our goal is to obtain a dense 3D reconstruction of the face without introducing any bias towards face models. The reconstruction will be used for forensic facial comparison.

In previous work [18], we presented a coarse 3D shape reconstruction method for faces. This coarse shape reconstruction method is model-free, which is important for usage in a forensic context. The reconstruction method is based on 20 manually

labelled landmarks in multiple views to support forensic data with low quality images. First one pair of images is selected for the start of an iterative reconstruction process. A 3D shape estimate is obtained based on the initial pair of images. In each iteration, one new image is added and both the 3D shape and the view of the face in each frame are optimised simultaneously. The reconstruction method is a structure from motion approach that minimises the 2D reprojection error, the error between the reprojected landmarks and the landmarks in the input frames. Several steps were taken to provide a robust reconstruction method for a low number of landmarks. The output of the 3D shape reconstruction algorithm is an estimate of the 3D points and an estimate of the rotation and translation parameters of the face in each view. The method is tested on simulated data: a random point clouds and a set of views rendered from a styrofoam head. An analysis on the 2D reprojection error and the 3D error using a varying number of views is provided, for more details we refer to [18]. Finally, our previous work shows that the shape reconstruction algorithm is capable of handling real image data.

In this paper, we propose a 3D reconstruction method based on a coarse 3D shape and a powerful texture reconstruction method to obtain dense 3D reconstructions and frontal reconstructions of high quality. The coarse 3D shape reconstruction is obtained using the shape reconstruction algorithm in [18], including the rotation and translation estimations of the views. The rotation and translation parameters for a view v define a rigid transformation T_v, that we use for illumination estimation. Since the reconstructed coarse shape only consists of 3D points, we define 28 triangular patches on the reconstructed coarse shape, see Fig. 1, to obtain a 3D surface. The patches are chosen in such a way that the surface has a seam running along the symmetry plane of the face. The patches are used for initialisation of the dense shape reconstruction.

The proposed dense reconstruction method is based on the Lambertian reflectance model, where we assume ambient light l_α and a single directional light source $\boldsymbol{l}_\mathrm{d}$. The light source $\boldsymbol{l}_\mathrm{d}$ is posed perpendicular to the frontal face. The illumination model we describe is object oriented and therefore we transform the constant directional light source $\boldsymbol{l}_\mathrm{d}$ to object coordinates and make the light direction dependent on view v. We use a rigid transformation T_v, based on the estimates of the rotation and translation parameters, to calculate the light direction in each view, see as follows:

$$\boldsymbol{l}_v = T_v(\boldsymbol{l}_\mathrm{d}) \tag{1}$$

We define our illumination model in the following equation:

$$\hat{G}_{xv} = \alpha_x l_\alpha + \rho_x(\boldsymbol{n}_x^\top \boldsymbol{l}_v) \tag{2}$$

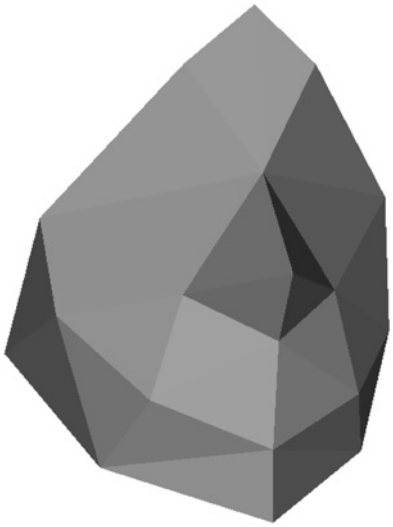

Fig. 1 *Coarse 3D shape reconstruction with added surface*

where α_x is the ambient reflection coefficient, l_α is the ambient light intensity, ρ_x is the diffuse reflection coefficient (albedo) in point x, $\boldsymbol{n}_x$ is the 3D normal in point x, $\boldsymbol{l}_v$ is the directional light source, including the light intensity, dependent on view v, see (1), and $\hat{G}_{xv}$ is the predicted intensity in point x using view v. The Lambertian reflectance model connects the 3D shape, via the normals of an object, to the observed intensities in images. We use this connection to optimise both the texture and the 3D shape. From the coarse reconstruction of the shape, estimates of the normals in each 3D point are available, but need to be optimised. Each 3D point x can be projected to multiple views, which we use to optimise the albedo and the normal of each point. We define the following optimisation, see as follows

$$(\rho_x^*, \boldsymbol{n}_x^*) = \arg\min_{\rho_x, \boldsymbol{n}_x} \sum_{v \in V} |G_{xv} - \hat{G}_{xv}| \tag{3}$$

where ρ_x^* is the optimised albedo, $\boldsymbol{n}_x^*$ is the optimised normal, V is the set of frames used for the reconstruction, G_{xv} is the observed intensity of point x in view v and $\hat{G}_{xv}$ is the predicted intensity given (2), where the term $\alpha_x l_\alpha$ is constant and has no influence on the optimisation. As there is an ambiguity between the intensity of $\boldsymbol{l}_v$ and the albedo ρ_x for optimisation purpose we assume the L2-norm of $\boldsymbol{l}_v$ to be one. The direction of the light in $\boldsymbol{l}_v$, however, is still depending on view v. The albedo ρ_x and the normals $\boldsymbol{n}_x$ of the 3D reconstruction are both optimised by minimisation of the absolute difference with the observed intensities. For each point x that is visible in at least four views, we are able to optimise the albedo and the normal. The normals are calculated from the shape for every point x using its neighbours, see Fig. 2, where x is the index of the central point and 1–4 are the indices of the neighbouring points. The 3D points are represented in a depth map, where each point in the grid contains the Z-coordinate as depth.

The normals of the four patches A–D are calculated with (4), where $\boldsymbol{p}_x$ is the central point, $\boldsymbol{p}_i$ are the neighbours counted clockwise, see Fig. 2, starting with the leftmost point and $\|\|_2$ denotes the Euclidean norm

$$\boldsymbol{n}_{xi} = \frac{(\boldsymbol{p}_x - \boldsymbol{p}_i) \times (\boldsymbol{p}_x - \boldsymbol{p}_{(i \bmod 4)+1})}{\|(\boldsymbol{p}_x - \boldsymbol{p}_i) \times (\boldsymbol{p}_x - \boldsymbol{p}_{(i \bmod 4)+1})\|_2} \tag{4}$$

The normal in point x is then calculated based on the triangular patches A–D, see (5), where we define the depth of $\boldsymbol{p}_x$ as: $d_x = [\boldsymbol{p}_x]_z$, the Z-coordinate of $\boldsymbol{p}_x$

$$\hat{\boldsymbol{n}}_x(d_x) = \frac{\boldsymbol{n}_{x1} + \boldsymbol{n}_{x2} + \boldsymbol{n}_{x3} + \boldsymbol{n}_{x4}}{\|\boldsymbol{n}_{x1} + \boldsymbol{n}_{x2} + \boldsymbol{n}_{x3} + \boldsymbol{n}_{x4}\|_2} \tag{5}$$

On the basis of the optimised normals we adapt the depth of the 3D shape (and the associated normals) to fit the optimised normals, see (6), where d_x^* is the optimised depth in point x

$$d_x^* = \arg\min_{d_x} \|\boldsymbol{n}_x^* - \hat{\boldsymbol{n}}_x(d_x)\|_1 \tag{6}$$

We perform the full process multiple times using different resolutions of the 3D shape to obtain our final reconstruction and

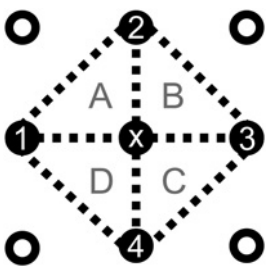

Fig. 2 *3D point x on a grid and its neighbours form the triangular patches A–D. These patches are used to calculate the normal in point x*

Algorithm 1

Input: V is a set of rotation and translation parameters
S is a coarse shape description with surface
I is a sequence of non-frontal face images

Output: G is a grid with depth values
A is a set of albedo values with the same resolution as G

1: **procedure** ITERATIVEDENSERECONSTRUCTION(V, S, I)
2: $(V, S) \leftarrow$ TransformToFrontalShape(V,S) ▷ Transformation, see Equation (1)
3: $G \leftarrow$ SampleShapeToGrid(S) ▷ Sample 3D depth values to grid
4: **while** Resolution(G) $\neq$ MaxResolution **do**
5: $G \leftarrow$ SmoothDepthGaussian(G) ▷ Smoothing needed due to upsampling
6: $N \leftarrow$ CalculateNormals(G) ▷ Set of normals, see Equation (5)
7: $(A, N) \leftarrow$ OptimiseNormalsAndAlbedo(V,N,I) ▷ Optimisation, see Equation (3)
8: $(G, N) \leftarrow$ AdaptDepthMatchNormals(G,N) ▷ Adapt the depth, see Equation (6)
9: $A \leftarrow$ RecalculateAlbedo(G,N,I) ▷ Calculate albedo using adapted depth
10: $(G, N) \leftarrow$ UpsampleGrid(G,N) ▷ Obtain 3D shape with higher resolution
11: **end while**
12: **return** G, A ▷ Return optimised albedo and dense 3D shape description
13: **end procedure**

Fig. 3 *Dense reconstruction algorithm*

albedo estimations. The proposed reconstruction method is described in Algorithm 1 (see Fig. 3).

The shape and the albedo of the face are both optimised and provide a more detailed and dense 3D shape description of the face. In each iteration of the algorithm, we obtain an estimate of the albedo of the face and a description of the 3D shape. Note that the proposed method does not optimise or improve the rotation and translation parameters and the internal camera calibration values of the input data, that are defined in V. These parameters are considered precise enough to perform the optimisation.

The coarse 3D shape is aligned in Step 2 using the eigenvectors of the symmetry plane of the reconstructed shape and by rotating the vector, defined by the tip of the nose and the centre of gravity of the symmetry plane, around the axis through the centre of gravity of the symmetry plane, to a frontal position. The face is now in an object oriented coordinate system based on the shape. In Step 3, a grid search is used to sample the nearest neighbours on the surface of the coarse 3D shape. A grid of 100×100 points is an appropriate starting point for the proposed method. The size of the grid is important, because the smaller the grid the more 3D information is described by each point in the grid. So, with a smaller size of the grid larger shape deformations occurs, for a large size of the grid only small details of the shape are changed. We stop at a size of 400×400 points, because the alterations are barely visible at this stage. The factor for upsampling in Step 10 in the algorithm is also of importance, because of the 3D information described by each point in the grid. The upsampled depth values of the grid are calculated by bilinear interpolation. We set the factor for upsampling to $\sqrt{2}$, which is appropriate according to the details that need to be reconstructed. Throughout an iteration only the depths of the points on the grid are adapted. In the optimisation in Step 7, we use a selection of normals with an angle smaller than 60° compared to the direction of the light, since these normals are more accurate than normals with larger angles. Although this angle could be smaller, this would lead to points in the grid where we are unable to minimise (3) due to the low number of normals. The minimisation of (3) is attained by finding the optimum values using an exhaustive search within small variations of the angle, at most 5° in both directions in multiple steps, for the estimated normals and by calculating the albedo values in 256 steps within the range [0…1]. During the shape adaptation in Step 8, we test three options for each point in the grid: we increase the depth, we decrease the depth or we keep the current depth. For each of these options, we calculate the absolute difference between the optimised normals in Step 7 and the normals calculated from the current 3D shape, see (6). We choose the depth with the smallest absolute difference. The altered depth is directly applied on each point. We continue altering the depth until there is no more change of depth. We used depth alterations of 0.5 mm at the time. The small steps lead to a slower convergence, but prevent extreme changes in the normals of the shape. Secondly, the small steps support a continuous 3D shape reconstruction.

In Fig. 6 on the right, an example of a frontal albedo reconstruction is shown. Fig. 11 shows multiple renderings of a reconstructed 3D shape with the reconstructed albedo as texture.

3 Experiments

To evaluate the performance of the proposed reconstruction method, we performed a 2D face comparison experiment on the albedo reconstructions. We compared the reconstructed albedo images with the ground truth images of the corresponding persons. Secondly, we evaluate the implications for forensic face recognition. In the last part of this section, we also visually inspect the reconstructed 3D models to indicate the quality of the 3D reconstruction.

3.1 Dataset

The dataset required for our experiments cannot be extracted from benchmark datasets such as CMU multi-PIE, because in those datasets the data is taken at the same moment in time with multiple cameras. For these recordings the illumination estimation and depth estimations become much easier, because of the constant brightness and lack of expression and deformation of the face. Therefore, we had no choice but to record our own dataset.

Fig. 4 *Examples of the dataset from multiple subjects with different views, showing minor expressions and some image artefacts*

Our dataset consists of recordings of 48 people. Each recording contains 101 frames with different views of the face of a person, recorded with a single camera. In the experiments we use subsets of this dataset, based on the availability of a set of frames suitable for reconstruction and a coarse shape estimate with a reasonable reprojection error. The selection criterion of the subsets is explained later in this section. Fig. 4 shows some samples of our dataset. The persons in the data set were asked to rotate their heads to the left and to the right, while looking up or down at the same time. As a result, a variety of frontal, near-frontal and frames under pose are captured in the dataset. The rotation and movement of the head simulate the actions performed at an ATM. All the frames were captured in about 45 s, similar to forensic time lapse recordings. Ground truth frontal views were taken afterwards with a different camera setup, during the same session.

3.2 2D face comparison

In the first experiment, we use the albedo reconstructions for comparison with ground truth images. The albedo reconstructions are stored in each iteration of the proposed method with increasing resolutions. Each albedo reconstruction is compared to the corresponding ground truth frontal view. We use the B7 algorithm of FaceVACS for face comparison, which is part of a commercial face comparison SDK [19]. FaceVACS is known in the face comparison community to be an excellent 2D face comparison algorithm. The comparison scores of FaceVACS are in the range [0…1]. Usually a threshold of 0.4 or above is taken for genuine

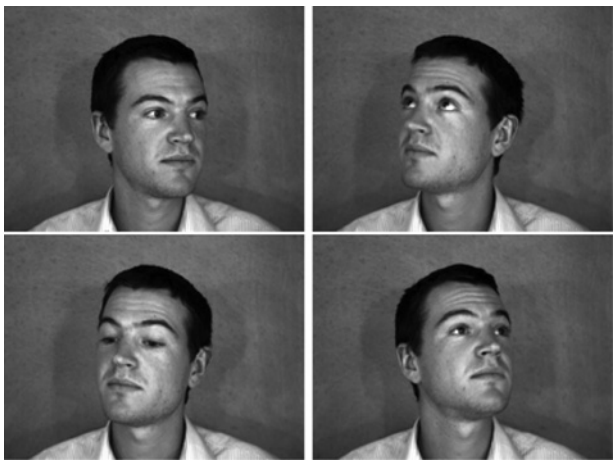

Fig. 5 *Examples of four frames from an input set of 30 frames*

Fig. 6 *Right: Example of reconstruction in the highest resolution. Left: corresponding gallery image*

scores. We use 0.5 in our experiment as genuine threshold. To show the ability of the proposed reconstruction method we select a subset of 2D frames from the dataset with FaceVACS scores below 0.5. These frames, which have non-frontal pose, expression or motion artefacts, are used in the reconstruction process, see Fig. 5 for some examples. We refer to this set as selection Σ. Improvement of the frames in selection Σ is important in forensic cases where no frontal images of the face are available. We used a subset of 30 frames from the selection Σ to ensure that every reconstruction is based on the same number of frames. This is also an appropriate number of frames for shape reconstruction according to earlier experiments, see [18], and for a reliable estimation of the albedo of the face using the optimisation in (3). The subset of 30 frames from selection Σ is selected based on their variation in view and image sharpness.

In Fig. 6, an example of one of the reconstructions is shown together with its corresponding ground truth image. In most of the cases the reconstructions seems to be slightly squint-eyed, because of the different origin of the views in the eyes. We notice that, although, humans are sensitive to the squint reconstruction artefacts, the performance of the FaceVACS algorithm is not affected by these artefacts.

The total result of the experiment is shown in Fig. 7. This graph shows the cumulative percentage of scores above a certain FaceVACS score for all reconstructions based on the selection Σ. For example, 20% of the reconstructions have a FaceVACS score above 0.8. The scores are the maximum scores taken over all albedo reconstructions of different resolutions of each person. The light grey area visualises the gain of the proposed reconstruction method. The grey area indicates the maximum scores of the input frames from selection Σ. The dark grey area indicates reconstruction scores below the threshold of 0.5. In that case taking the best input frame would probably give better results. The graph shows that in 91% of the reconstructions we surpassed the threshold score 0.5 of the input set of frames. In quite some cases, we have a considerable gain, in some cases even up to 0.99. Such scores are in the same range as high quality frontal face images. Only in 9% of the reconstructions the proposed method did not improve the recognition results.

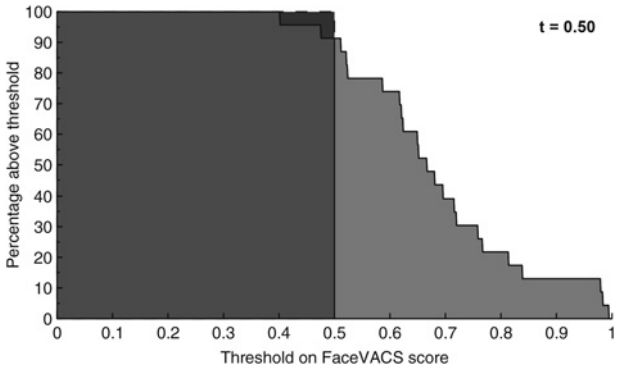

Fig. 7 *FaceVACS reconstruction results exceeding the threshold in 91% of the cases*

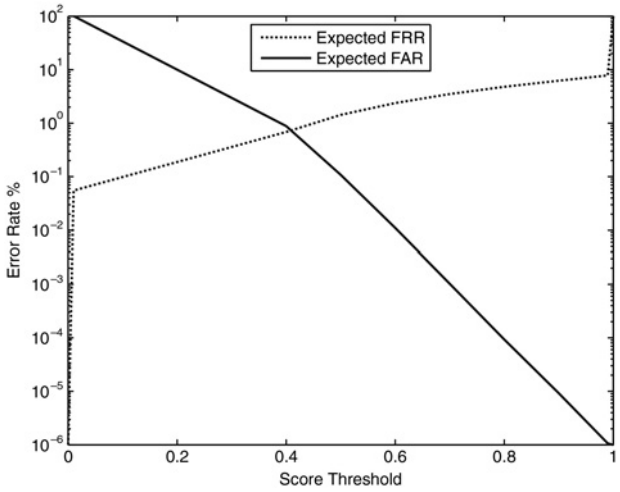

Fig. 8 *Expected FAR and FRR from FaceVACS algorithm*

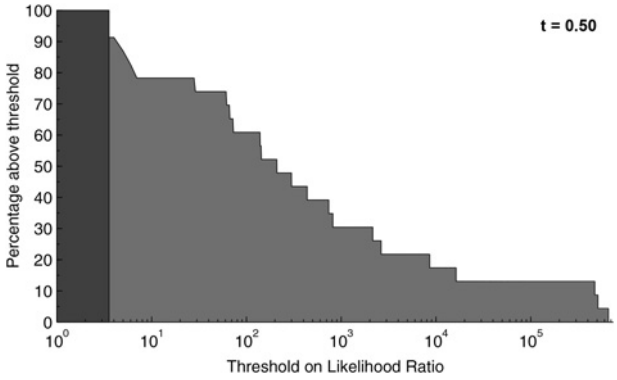

Fig. 9 *Likelihood ratios of the reconstructions. Grey area corresponds to the maximum-likelihood ratio of the frames in selection Σ for each person. Light grey area shows the likelihood ratio of the reconstructions using selection Σ as input frames*

3.3 Implications for forensic face comparison

To give an indication of how FaceVACS scores should be interpreted in a forensic context, we provide likelihood ratios based on the statistics in the FaceVACS SDK [19]. Here, likelihood ratios describe the ratio between the probability of a comparison score, given that that faces compared originate from the same person and the probability of a comparison score, given that that faces compared originate from different persons, see [20]. This is the standard approach when evaluating the meaningfulness of forensic evidence. If we calculate the likelihood ratios for all FaceVACS scores, we can compare the result of the best 2D input frame with the reconstruction to see how the reconstruction increases the likelihood ratio. The FaceVACS scores can be converted into likelihood ratios based on the false acceptance rate (FAR) and false recognition rate (FRR), derived from a large test database [19]. In Fig. 8, a recreated FRR and FAR graph is shown. The likelihood ratios can be calculated from this graph by applying (7) to each score s, see [21]

$$\mathrm{LR}(s) = \frac{1 - (\partial(\mathrm{FRR})/\partial s)}{(\partial(\mathrm{FAR})/\partial s)} \tag{7}$$

The statistics provide a good indication of the likelihood ratios of the FaceVACS scores. As can be seen in Table 1 the likelihood ratio can increase by a factor of up to 1.0×10^5 due to the reconstructed albedo images. In more than 75% of the reconstructions the likelihood ratio increases by a factor of 10 or higher, which is a considerable gain. Since the statistics in Fig. 8 are valid for both the genuine and non-genuine comparisons, there is no additional gain for the non-genuine reconstructions.

If we convert the FaceVACS scores of the results to likelihood ratios, the reconstructions can be reviewed in a forensic context, see Fig. 9. The maximum-likelihood ratio for each set of input frames of selection Σ is 3.5, based on the score threshold of 0.5. If a forensic researcher takes, for example a likelihood ratio of 100 as a minimum, the proposed method is able to provide meaningful results in more than 60% of the cases.

Table 1 FaceVACS likelihood ratio per score of the FaceVACS algorithm

Score	Likelihood ratio
0.4	0.4
0.5	3.5
0.6	4.0×10^1
0.7	4.9×10^2
0.8	6.2×10^3
0.9	7.2×10^4
1.0	7.5×10^5

In Fig. 10, on the top we show the results for all 23 reconstructions individually. The same threshold of 0.5 was used for the selection of the input frames. For the other 25 persons in the dataset there was not enough low quality data available to obtain appropriate reconstructions. Increasing the quality threshold of 0.5 for the frame selection would enable the availability of more and higher quality frames in the selection, however, this would be a less challenging problem. Continuing the reconstruction procedure with less than 30 frames, sometimes even less than 10, would lead to less accurate frontal face reconstructions. The light grey bars in Fig. 10 indicate that we were able to exceed the threshold scores for that current person. The grey bars indicate the opposite. We are able to surpass the threshold score for 21 persons in the dataset.

In some occasions, the reconstruction scores can be further improved by changing the field of view of the frontal view. We included an experiment where we changed the field of view of the frontal view to 30°, which approximately fits the field of view of the camera of the input images. This was done by rendering the 3D shape including the albedo texturing of the 400×400 grid with a fixed field of view of 30° to a frontal view. In some cases, this gave an improvement of the results, see the bottom of Fig. 10, but in other cases the performance was worse. We can take the maximum FaceVACS score over both experiments to get the optimal result. In some rare occasions, there are strong motion artefacts and the proposed method is not able to surpass the FaceVACS scores. The only way to improve the scores in such a case would be to manually select the set of input frames to minimise motion artefacts and other abnormalities. We decided to not further pre-process or post-process the results to give a clear demonstration of the performance of the proposed algorithm.

We expect the quality of the albedo images to increase along with the resolution. If we look into the scores for all resolutions we notice that the bigger the resolution of the grid the higher the percentage which surpasses the comparison threshold, see Table 2. The gain here represents the average increase over all 23 reconstructions of the FaceVACS scores compared to the threshold score. The maximum gain possible is 0.5. We decided to stop at 400×400 points for the grid. Although higher resolution reconstructions still have a positive effect on the gain, the calculation time increases quadratically and the gain becomes smaller each time. The last row shows the average increase for the resolution with maximum score, because sometimes an increase of resolution does not result in an increase of the FaceVACS scores.

3.4 3D visual inspection

Apart from the albedo, also the 3D shape is reconstructed with the proposed algorithm. Some examples of these full 3D reconstructions can be seen in Fig. 11. The 3D models look quite

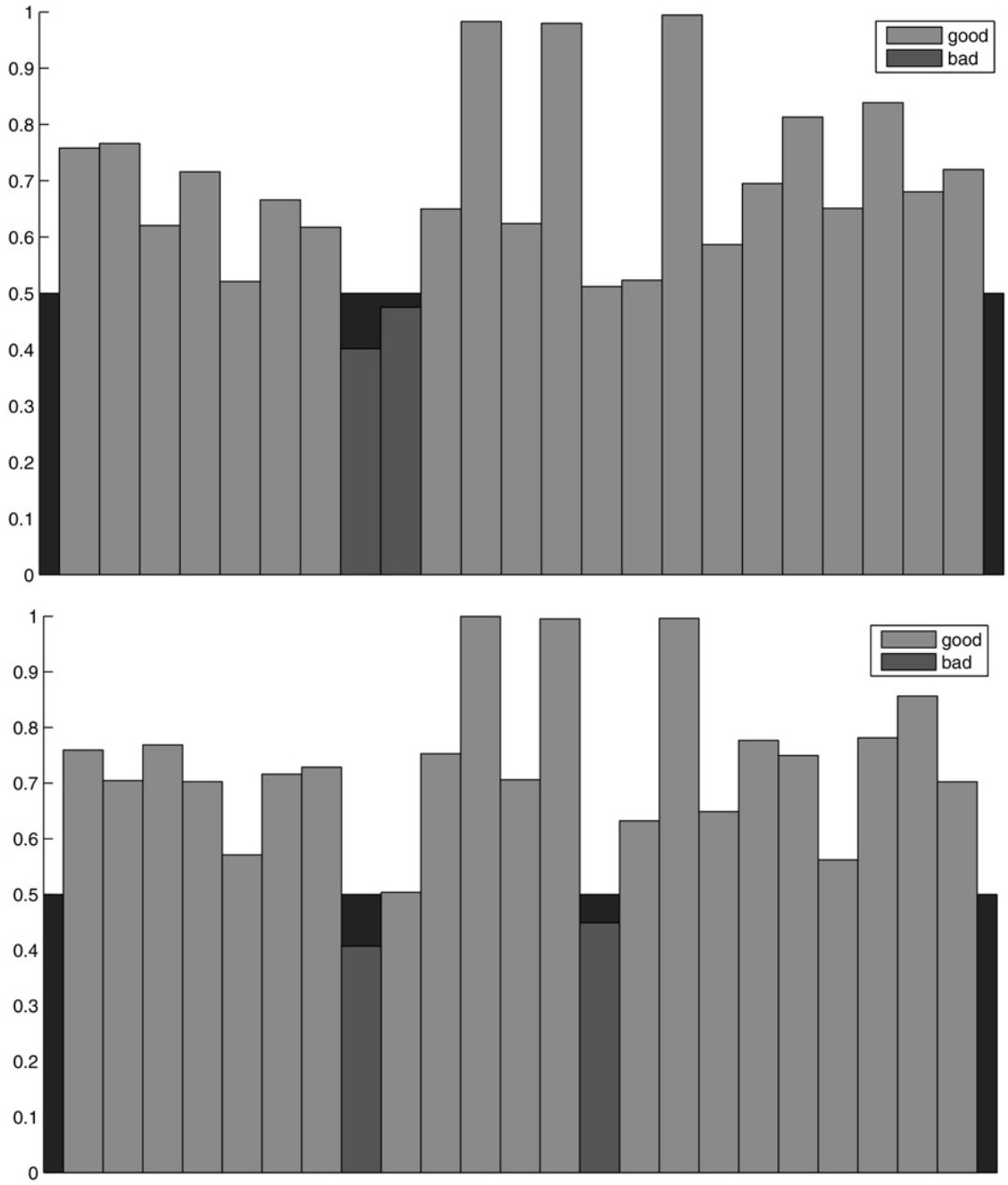

Fig. 10 *Top: FaceVACS scores per person. Bottom: FaceVACS scores per person using a fixed field of view of 30°*

realistic, but in most cases the nose is a bit flattened by the smoothing in the algorithm. For frontal views this effect is minimal, but for views under pose the effect is stronger. Small details of the 3D shape, like the shape of the lips and the shape of the eyes are visible on the reconstructed 3D models. Although, the 3D shape seems quite accurate, experiments with shape only comparison showed that the shape is not near the quality of laser scans and structured light models. The reconstructed 3D models are of high enough quality to correct the pose of the face, but cannot be used for shape comparison.

Table 2 FaceVACS recognition score gain for multiple reconstruction sizes

Size	Gain
100 × 100	0.04
141 × 141	0.10
200 × 200	0.11
283 × 283	0.15
400 × 400	0.17
maximum	0.19

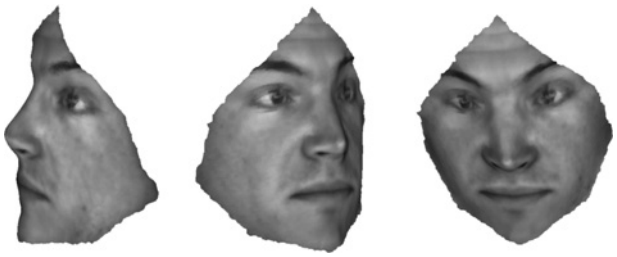

Fig. 11 *Multiple renderings of a reconstructed 3D face shape*

4 Conclusion

We explored the possibilities of a dense model-free 3D face reconstruction method, based on image sequences from a single camera, to improve the current state of forensic face comparison. The forensic context implies that models based on facial data cannot be used, because they cause a bias towards those faces. We proposed a dense 3D reconstruction method with two stages. In the first stage, we obtain a coarse 3D landmark shape and an estimate of the rotations and translation in each frame. In the second stage, we use the Lambertian reflectance model to estimate the albedo and to refine the 3D shape of the face. In our experiments, the proposed multi-resolution approach is able to deliver frontal face reconstructions that improve the face comparison scores in more than 90% of the cases. Therefore, we reached our goal of improving face recognition performance without introducing any bias towards a face model. Using the likelihood ratio framework, we show for the same experiment that for data initially unsuitable for forensic use, the reconstructions become meaningful in a forensic context in more than 60% of the cases. Visual inspection of the 3D shape shows that the reconstructed 3D shape with albedo texture can be used to generate faces under pose, but not for shape-based 3D face recognition.

5 Acknowledgment

This work was supported by BZK 5.50: 'Gezichtsvergelijking op basis van niet gekalibreerde Camerabeelden' in a cooperation between the Netherlands Forensic Institute and University of Twente. The authors thank Cognitec Systems GmbH for supporting our research by providing the FaceVACS software. Results obtained for FaceVACS were produced in experiments conducted by the University of Twente, and should therefore not be construed as a vendor's maximum effort full capability result.

6 References

1 Blanz, V., Vetter, T.: 'A morphable model for the synthesis of 3d faces'. Proc. of the 26th Annual Conf. on Computer Graphics and Interactive Techniques, 1999, pp. 187–194
2 Fua, P.: 'Regularized bundle-adjustment to model heads from image sequences without calibration data', *Int. J. Comput. Vis.*, 2000, **38**, (2), pp. 153–171
3 Shan, Y., Liu, Z., Zhang, Z.: 'Model-based bundle adjustment with application to face modeling'. IEEE Int. Conf. on Computer Vision, 2001, vol. 2, p. 644
4 Roy-Chowdhury, A.K.: '3d face reconstruction from video using a generic model'. Int. Conf. on Multimedia and Expo, 2002, pp. 449–452
5 Kang, S.B., Jones, M.: 'Appearance-based structure from motion using linear classes of 3-d models', *Int. J. Comput. Vis.*, 2002, **49**, (1), pp. 5–22
6 Chowdhury, A.K.R., Chellappa, R.: 'Face reconstruction from monocular video using uncertainty analysis and a generic model', *Comput. Vis. Image Underst.*, 2003, **91**, pp. 188–213, special Issue on Face Recognition
7 Roy-Chowdhury, A.K., Chellappa, R., Gupta, H.: '3D face modeling from monocular video sequences' (Academic Press, 2005), Ch. 6, pp. 185–218
8 Fidaleo, D., Medioni, G.G.: 'Model-assisted 3d face reconstruction from video', in Zhou, S.K., Zhao, W., Tang, X., Gong, S. (Eds.): 'AMFG', ser. Lecture Notes in Computer Science (Springer, 2007), vol. 4778, pp. 124–138
9 Park, U., Jain, A.K.: '3d model-based face recognition in video', in Lee, S.-W., Li, S. (Eds.): 'Advances in biometrics', ser. Lecture Notes in Computer Science (Springer Berlin Heidelberg, 2007), vol. 4642, pp. 1085–1094
10 Marques, M., Costeira, J.: '3d face recognition from multiple images: A shape-from-motion approach'. FG, 2008, pp. 1–6
11 Torresani, L., Hertzmann, A., Bregler, C.: 'Nonrigid structure-from-motion: estimating shape and motion with hierarchical priors', *IEEE Trans. Pattern Anal. Mach. Intell.*, 2008, **30**, (5), pp. 878–892
12 Ishimoto, M., Chen, Y.-W.: 'Pose-robust face recognition based on 3d shape reconstruction'. Fifth Int. Conf. on Natural Computation, 2009, ICNC'09, 2009, vol. 6, pp. 40–43
13 Hamsici, O.C., Gotardo, P.F.U., Martinez, A.M.: 'Learning spatially-smooth mappings in non-rigid structure from motion'. Proc. of the 12th European Conf. on Computer Vision, ECCV'12, 2012, pp. 260–273
14 Spreeuwers, L.: 'Multi-view passive 3d face acquisition device'. Proc. of the Special Interest Group on Biometrics and Electronic Signatures, ser. Lecture Notes in Informatics (LNI) – Proc., September 2008, vol. P-137, pp. 13–24
15 Strecha, C., Fransens, R., Van Gool, L.: 'Combined depth and outlier estimation in multiview stereo'. IEEE Computer Society Conf. on Computer Vision and Pattern Recognition, 2006, 2006, vol. 2, pp. 2394–2401
16 Garg, R., Roussos, A., Agapito, L.: 'Dense variational reconstruction of non-rigid surfaces from monocular video'. IEEE Conf. Computer Vision and Pattern Recognition (CVPR), 2013, June 2013, pp. 1272–1279
17 Delaunoy, A., Pollefeys, M.: 'Photometric bundle adjustment for dense multi-view 3d modeling'. IEEE Conf. Computer Vision and Pattern Recognition (CVPR), 2014, June 2014, pp. 1486–1493
18 van Dam, C., Spreeuwers, L., Veldhuis, R.: 'Landmark-based model-free 3d face shape reconstruction from video sequences'. Proc. of the Int. Conf. of Biometrics Special Interest Group 2013 BIOSIG, September 2013, pp. 265–272
19 Cognitec Systems GmbH: 'FaceVACS SDK 8.8.0', http://www.cognitec-systems.de, 2013
20 Meuwly, D., Veldhuis, R.N.J.: 'Biometrics – developments and potential', in Allan, Jamieson, Andre, A. Moenssens (Eds.): 'Wiley encyclopaedia of forensic science' (John Wiley & Sons Ltd, Chichester, UK, 2014), vol. 1, pp. 1–8
21 Veldhuis, R.: 'The relation between the secrecy rate of biometric template protection and biometric recognition performance'. Int. Conf. on Biometrics (ICB), 2015, 2015, vol. 5, pp. 311–318