Vision, Modeling, and Visualization (2016)
M. Hullin, M. Stamminger, and T. Weinkauf (Eds.)

3D Face Reconstruction with Silhouette Constraints

Q. Hu M. Zwicker and P. Favaro

Institute of Computer Science, University of Bern, Switzerland

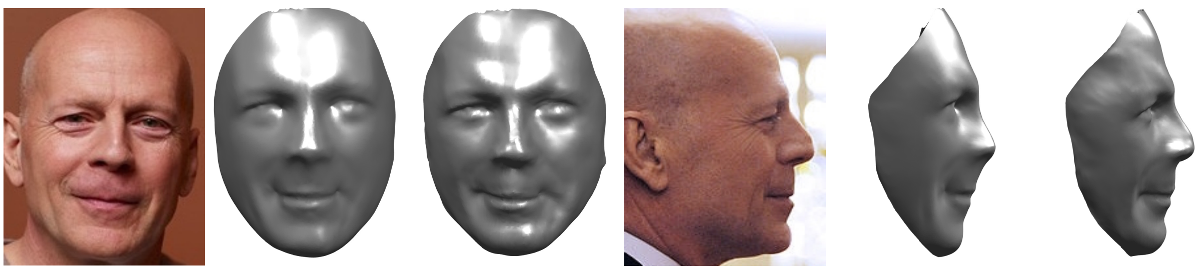

Figure 1: *We introduce silhouette constraints to improve the quality of unconstrained 3D face reconstruction. Here, we show a comparison between the state of the art approach by Roth et al. [RTL15] and our technique. The 2nd and 5th column are the results from [RTL15], the 3rd and 6th column are our results. Note how our approach follows more faithfully the silhouettes of the input images, especially around the nose region.*

Abstract

In this paper we introduce silhouette constraints to improve the quality of unconstrained 3D face reconstruction. Previously, state of the art unconstrained 3D face reconstruction techniques relied solely on photometric consistency and matching sparse facial landmarks. We show that constraining the silhouettes of the 3D reconstruction to match silhouettes in the input images can further improve reconstruction quality. Our technique automatically detects silhouettes and iteratively matches silhouette points computed from the current 3D reconstruction with silhouettes in the input images. We demonstrate that our results improve on the previous state of the art in unconstrained 3D face reconstruction and that our additional constraints can easily be included in the iterative reconstruction at little additional cost.

1. Introduction

We consider the problem of 3D face reconstruction from internet photo collections. Our goal is to reconstruct 3D models of individuals from collections of images in uncontrolled environments, including variations in illumination, pose, and expression, which has been called "face reconstruction in the wild" [KSS11] or "unconstrained 3D face reconstruction" [RTL15]. Such 3D reconstructions can be useful for face and expression recognition [LC05, WYWS06], or to produce facial animations [LYYB13].

Our work is inspired by recent progress in unconstrained face recognition by Roth et al. [RTL15]. They leverage state of the art photometric stereo techniques, recent advances in 2D facial landmark estimation, and a full 3D face representation (instead of 2.5D height fields) to obtain impressive results, considering that the input consists of images under various illumination conditions, with different poses and facial expressions, and neither video or stereo data is included in the input. Nonetheless, the quality of the 3D reconstructions is limited since the only constraints on the reconstruction are photometric consistency and correspondence with sparse facial landmarks.

In this paper, we introduce silhouette constraints to improve the quality of unconstrained 3D face reconstruction. Our main idea is to extract silhouette points on the 3D reconstruction, and match them with automatically detected silhouette points in the input images. We include these constraints in the 3D reconstruction objective, which we solve in an iterative process. In each iteration step, we recompute the silhouette points using the current 3D reconstruction and update the corresponding contraints in the objective. As a consequence, the silhouettes of the 3D reconstruction converge towards the silhouettes in the input images. Our results demonstrate that the new silhouette constraints lead to higher reconstruction quality.

The rest of this paper is organized as follows: We first review related work in Section 2, and provide a brief overview of state of the art unconstrained face reconstruction as proposed by Roth et al. [RTL15] in Section 3. In Section 4 we introduce the novel silhouette constraints as our main contribution. Finally, we present results in Section 5 and conclude in Section 6.

2. Related work

Face Reconstruction from Image Collections. Face reconstruction "in the wild" from collections of images under uncontrolled illumination, and with varying facial pose and expressions, has been a long standing problem in computer vision. State of the art methods are mostly based on photometric stereo, such as the pioneering work by Kemelmacher-Shlizerman and Seitz [KSS11], and its extension to video input [SKSS14]. This approach has been improved recently by Roth et al. [RTL15] who solve for a full 3D mesh instead of a 2.5D height field, and leverage state of the art 2D facial landmark estimation [YLYL13a]. We build on the work by Roth et al., but extend it to include silhouette constraints to overcome some of the limitations of photometric stereo.

Face Tracking and Animation from Video. Research in the Computer Graphics community has achieved impressive results for the problem of tracking, reconstructing, and animating faces based on video data. Generally, state of the art techniques represent faces and facial expressions using dynamic expression models, also called blendshape models [CWZ*14]. While earlier work relied on RGBD video input [WBLP11, LYYB13, BWP13], it is now possible to solve the tracking and reconstruction problem based on RGB video data in real-time at impressive quality. [CHZ14, CBZB15, GZC*16, TZS*16]. The key idea in this work is to learn a regression model that maps 2D video frames to 3D facial landmarks, and then to register the DEM to the 3D landmarks [CWLZ13]. While this approach required calibration for individual users, it has then been extended to a calibration-free approach that iteratively refines the model [CHZ14]. A most recent extension also synthesizes highly detailed geometry such as wrinkles [CBZB15]. The key difference to our work is that these techniques either require calibration for each user, or they rely on the coherence of video input to adapt the model in an iterative manner.

3. Unconstrained Face Reconstruction

Unconstrained 3D face reconstruction [RTL15] takes as its input a collection of facial photographs of an individual under uncrontrolled illumination and various poses and facial expresssions, and it returns a full 3D reconstruction of the individual's face represented by a mesh. The method proceeds by first detecting 2D landmarks on all input images using the approach by Yan et al. [YLYL13b]. Next, a 3D template mesh is warped and projected to 2D to match the 2D landmarks. This leads to an initial, rough 3D reconstruction of the face, and a weak perspective projection matrix for each input image. Then they improve the 3D reconstruction by determining surface normals using a photometric stereo approach, and lastly they perform 3D reconstruction by refining the rough mesh to match the photometric normals. We briefly review the main components of this approach.

Template Warping. Given the 2D landmarks, denoted by a vector W_i, for all images i, and a template mesh with p vertices, denoted by a $3p$ dimensional vector X_0, the goal is to find a warped mesh X and weak projection matrices P_i, such that the landmarks on X projected to 2D match the 2D landmarks W_i. The key idea is to use Laplacian surface editing [SCOL*04] to perform mesh deformation. This leads to a minimization problem

$$E_{warp}(X, P_i) = \|\mathcal{L}X - \mathcal{L}X_0\|^2 + \lambda_l \sum_i \|P_i D_i X - W_i\|^2, \quad (1)$$

which expresses the intuition that the mesh Laplacian [MDSB03] $\mathcal{L}X$ of the deformed mesh should stay close to the mesh Laplacian of the template $\mathcal{L}X_0$. In addition, D_i is a selection matrix picking out the landmarks that have a correspondence in image i, that is, it is a diagonal matrix with 1 on the diagonal for the vertices corresponding to a landmark and 0 everywhere else.

Equation 1 is solved iteratively. First, the projection matrices are obtained by fixing the template and solving for P_i, which is a linear least squares problem. Then, the P_i are fixed and the deformed mesh X is obtained in an iterative approach. This is necessary because the mesh Laplacians $\mathcal{L}X$ and $\mathcal{L}X_0$ are not rotation invariant. Hence, after each iteration step the mesh Laplacian $\mathcal{L}X_0$ of the template is rotated until it aligns with the Laplacian $\mathcal{L}X$ obtained in the previous step.

Photometric Normals. Next, a photometric stereo technique [KSS11] is used to estimate per vertex normals on the deformed template. First, an image reflectance intensity matrix $M \in \mathbb{R}^{n\times p}$ is constructed, where n is the number of images and p the number of mesh vertices. Each element i, j stores the reflectance intensity of the pixel in image i that corresponds to vertex j on the mesh, where correspondence is established by projecting the mesh onto image i using the projection matrix P_i. For non-frontal images some vertices are not visible, which leads to undefined matrix entries that are filled using matrix completion [ZLM09]. Matrix M is then factorized using SVD, and the rank-4 approximation estimates a light matrix $L \in \mathbb{R}^{n\times 4}$ and shape matrix $S \in \mathbb{R}^{4\times p}$, which represents the normals and albedos at each mesh vertex.

The bas-relief ambiguity, that is, $LS = (\tilde{L}A^{-1})(A\tilde{S})$ with ambiguous factors $\tilde{L}$ and $\tilde{A}$ and any invertible $A \in \mathbb{R}^{4\times 4}$, is resolved using the approach by Kemelmacher-Shlizerman and Seitz [KSS11]. First, the images that are modeled well by the rank-4 approximation, that is, such that $\|M - LS\| < \varepsilon$, are selected. Then, the ambiguity is recovered by solving for $\arg\min_A \|S^t - A\tilde{S}\|^2$, where S^t is the shape vector of the template. After estimating the albedo ρ, normals n are obtained.

Final Reconstruction. To leverage the vertex normals for final 3D reconstruction, the idea is to exploit the fact that the vertex Laplacians correspond to the vertex normals scaled by the mean curvature at each vertex [MDSB03]. Therefore, the shape (that is, vertex positions) X can be reconstructed from the normals n by minimizing $\|\mathcal{L}X - Hn\|^2$, where H is a diagonal matrix storing the mean curvatures at each vertex. The mean curvature H_i at vertex i can be estimated based on the vertex normals as

$H_i = \frac{1}{4A_i}\sum_{j\in N(i)}(cot\alpha_{ij} + cot\beta_{ij}))e_{ij}\cdot(n_i - n_j)$, where $N(i)$ is the set of incident neighboring vertices of vertex i, A_i is the sum of the triangles' areas incident to i, e_{ij} is the edge from i to j [MDSB03].

Note that the mean curvature formula degenerates into a 1D version on the boundary. This leads to separate constraints on the boundaries formulated as $\|\mathcal{L}_b X - \mathcal{K} b\|^2$, where $\mathcal{L}_{b,ij} = \frac{1}{e_{ij}}$, $\mathcal{K}$ is the geodesic curvature along the boundary, and b is the cross product between the surface normal and the boundary tangent.

The final mesh is reconstructed by minimizing the energy that includes the Laplacian constraints on the interior and the boundary and the 2D landmark constraints

$$E = \|\mathcal{L}X - H^k n\|^2 + \lambda_b\|\mathcal{L}_b X - \mathcal{L}X^k\|^2 + \lambda_l \sum_i \|P_i D_i X - W_i\|^2. \tag{2}$$

Fixing the projection matrices P_i makes this a linear least squares problem for X. It is solved iteratively, however, by updating the cotan weights to compute the mean curvatures H after each step. Finally, Roth et al. [RTL15] incorporate a heuristics to deal with shadowed regions.

While this approach leads to impressive results, we can see that the reconstructed profile often does not fit well with the input images, especially around the nose and chin. For example in Figure 1 the nose is supposed to be taller and the chin should be more curved. We address these limitations by proposing novel silhouette constraints to refine the reconstructed face shape, as discussed next.

4. Including Silhouette Constraints

The main idea in our work is to include additional silhouette constraints to obtain a better match between the 3D shape and the input images, and by doing so improve the 3D reconstruction. We achieve this by automatically extracting silhouettes both in the 2D images, and on the 3D model given the projection onto each image. We then build correspondences between the silhouette points in each image and on the 3D model (under the projection to each image), and finally incorporate these silhouette constraints in the final 3D reconstruction step. The main steps of our approach proceed as follows:

1. Reconstruct a rough 3D model by deforming a template and estimating initial perspective projection and rotation matrices to match 2D landmarks, as in the method by Roth et al. [RTL15].
2. Use photometric stereo to estimate per-vertex normals and mean curvatures as proposed by Roth et al. [RTL15].
3. Using the current projection and rotation matrices, extract 3D and 2D silhouette candidates for each image. Build up correspondences between 2D and 3D silhouette candidates. Discard silhouette candidate points that only show up in few images.
4. Reconstruct the face model including the silhouette constraints. Re-estimate the perspective projection and rotation matrices based on the updated reconstruction.
5. Go back to step 2 until convergence.

In the following subsections, we will give the details of steps 2 to 4.

4.1. Silhouette Extraction

To extract silhouettes on the 3D model corresponding to each input image, we first detect the points on the 3D mesh whose normals are parallel to the image plane. Given the estimated perspective projection matrix for the image, we can also estimate the rotation matrix R_i for the image. The view direction can then be estimated from the rotation matrix. Suppose the direction perpendicular to the frontal face is the z-axis, then for the i-th image, the view direction is $v_i = R_i[0,0,1]^T$.

Silhouette candidate points on the 3D model are those vertices whose normals are perpendicular to the view direction, that is, the cosine of the angle between the view direction and the vertex normal should be near zero, $\frac{|v_i\cdot n_j|}{\|v_i\|\cdot\|n_j\|} < \varepsilon$, where n_j is the normal of vertex j on the 3D mesh and ε is a small value near zero. To avoid noise, we consider as silhouette candidates only those vertices whose sets of incident faces include faces that are front-facing and faces that are back-facing to the view direction. Among the silhouette candidates, those vertices that are occluded by other parts of the face, for example the nose, have no corresponding edge on the 2D image. Hence we also discard these points. Occlusions can easily be detected by ray tracing or rendering the mesh using z-buffering, for example. The points satisfying the above constraints are considered as silhouette candidates on the 3D model, which we denote as X_{sil3D}.

To guarantee proper extraction of silhouettes from the 2D input images, we use only the "nonfrontal" images. We estimate the yaw, pitch and roll of the face pose for each image from the rotation matrix, and we select only those images in which the yaw of the face is bigger than a threshold. We use a Canny edge detector [Can86] to identify candidate silhouette points in the 2D images.

4.2. 3D-2D Silhouette Correspondences

Next, we describe how we establish correspondences between 3D silhouette points on the mesh, and silhouette candidates on the 2D images. Let X^i_{sil3D} denote the silhouette candidates on the 3D mesh for image i. We define this vector via the diagonal matrix Δ_i, which selects the 3D vertices from the shape X

$$X^i_{sil3D} = \Delta_i X. \tag{3}$$

Then, we project these points onto the corresponding 2D image using the current estimation of the projection and rotations matrices. Finally, for each 3D silhouette point projected onto the image, we establish the correspondence to the closest 2D point in the detected 2D silhouette X^i_{sil2D}. The vector B_i collects all the 2D points in X^i_{sil2D} corresponding to each 3D point in X^i_{sil3D} (the correspondence is bijective). The k-th 2D point in B_i is therefore defined as the point $X^i_{sil2D,j}$ such that

$$j = \arg\min_l \|P_i X^i_{sil3D,k} - X^i_{sil2D,l}\|^2. \tag{4}$$

Since the reconstructed model is close to the real shape of the individual, the projected silhouette $P_i X^i_{sil3D,k}$ should not be too far away from the corresponding edge. In addition, we set a threshold for the square distance, and if the distance is larger than the threshold, that point will be considered as outlier and rejected. This can also reduce the artifacts caused by images containing extreme facial expressions.

In our approach, the face model is reconstructed from the average normals of all images. The silhouette points that only appear in a few images, however, may result in unreliable 3D mesh reconstructions. This is particularly evident with strong facial expressions, such as an open mouth. For example, if we extract silhouettes from a laughing face, the detected curves of the cheek and chin are very different from those detected on the face at rest and match poorly to the silhouettes on the average 3D model. In our implementation, the silhouette points appearing in fewer than four images are discarded.

4.3. Reconstruction using Silhouette Constraints

Given the silhouette points on the 3D model and the corresponding silhouettes in 2D images, we add the silhouette constraints to the original energy from Equation 2 and obtain an extended energy

$$\begin{aligned} E = &\|LX - H^k n\|^2 + \lambda_l \sum_i \|P_i D_i X - W_i\|^2 \\ &+ \lambda_b \|L_b X - LX\|^2 + \lambda_s \sum_{i \in I_{sil}} \|P_i \Delta_i X - B_i\|^2 \end{aligned} \tag{5}$$

where I_{sil} is the set of indices of the images containing valid silhouette points, Δ_i is the diagonal matrix selecting the 3D vertices appearing on the silhouette in image i, and B_i stores the corresponding 2D silhouette points in image i as described in Section 4.2. Finally, λ_s is a silhouette constraint weight. While keeping the projection matrices P_i fixed, we find X by solving the linear system

$$\begin{aligned} &(\mathcal{L}^2 + \lambda_b \mathcal{L}_b^2 + \lambda_l \sum_i D_i P_i^T P_i D_i + \lambda_s \sum_{i \in I_{sil}} \Delta_i P_i^T P_i \Delta_i) X \\ &= \mathcal{L}H + \lambda_b \mathcal{L}_b^2 X + \lambda_l \sum_i (P_i^T) W_i + \lambda_s \sum_{i \in I_{sil}} P_i^T \Delta_i B_i. \end{aligned} \tag{6}$$

5. Results

In this section we present visualizations and results of our algorithm, and demonstrate the improvements over the previous state of the art method by Roth et al. [RTL15].

Dataset. As input data for our approach we collected photos of celebrities from various websites. First, we used the Google API to collect photos by searching for celebrities. However, only less than half of the results could be used for our experiments. Many of the returned images are not portrait photos or are photos without the person of interest. Hence, we collected additional images by using also the Bing API, websites Mtime and Douban with a python script to download images for different individuals. For the initial template used to reconstruct the model, we use the face model from Zhang et al. [LZS04]. The landmarks are detected using the method by Kazemi and Sullivan [KS14].

We reconstructed 3D face models of various celebrities, including George Clooney (1153 photos), Bruce Willis (834), Edward Norton (706), Tom Cruise (710), WentworthMiller(776), Colin Firth (824), James McAvoy (1125), and Tom Hanks (862). The photos which are used to extract silhouettes take up around 10% of the total amount for each individual. All photos were collected

Figure 2: *The silhouette on the 3D mesh projected to the image (red points) in iteration 1, 4 and 7 of our main iteration loop (Section 4). The silhouette gets closer and closer to the real silhouette as the iteration progresses. We update the silhouette points on the 3D mesh in each iteration.*

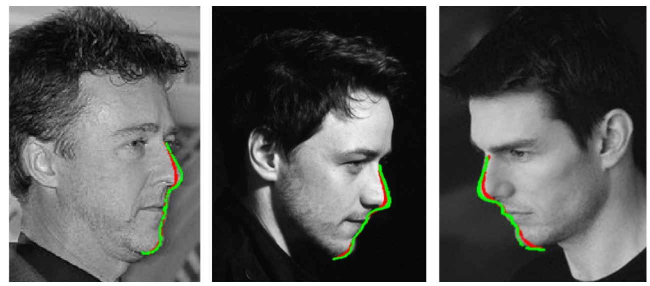

Figure 3: *We compare 3D silhouette points projected onto input images that were obtained after convergence with our silhouette constraints (green points) and without (red points) for three different individuals. Our approach clearly leads to a more faithful reconstruction of the true silhouettes in the input images.*

from the internet as mentioned above. We use OpenCV to crop the images to keep the face only. We do not scale the image size, and retain the resolution of the downloaded (cropped) image. The photos are completely unconstrained, including unknown illumination, pose, and facial expressions. Figure 6 shows a comparison of our results and the results from our implementation of the method by Roth et al. [RTL15]. In the profile views, we can see that our results are more accurate around the nose, chin, and around the eyes. In our method we observe a higher level of detail and less oversmoothing, because the silhouette constraints lead to better correspondences between the 3D mesh and the input images. In turn, these correspondences help estimate more accurate photometric normals.

Silhouette Constraint Visualization. In Figure 2 we visualize the silhouette on the 3D mesh by projecting the 3D silhouette points onto an input image as our main iteration loop (Section 4) progresses. We see how the projected silhouette points approximate the silhouette in the input image better and better. Figure 3 compares the silhouettes of the 3D reconstruction obtained with (green points) and without (red points) the silhouette constraints after convergence for three different individuals. Our silhouette constraints lead to more accurate correspondences with the true silhouettes in the input images.

Ground-truth Comparison. Figure 4 shows a comparison of our approach with silhouette constraints to the technique by Roth et al. [RTL15] with respect to a ground truth 3D geometry obtained using an Artec Eva 3D scanner. We use 807 images of the volunteer. We align the reconstructed faces with the ground truth using the landmarks on the two meshes. Then, we compute the Euclidean distance from each vertex on the ground truth to the closest point on the reconstructed mesh and normalize the distances. Figure 4 shows a color coded visualization of the distances, where red corresponds to big distances to the ground truth, and green represents small distances. We observe that our result matches better the ground truth especially in the area around the nose and the chin.

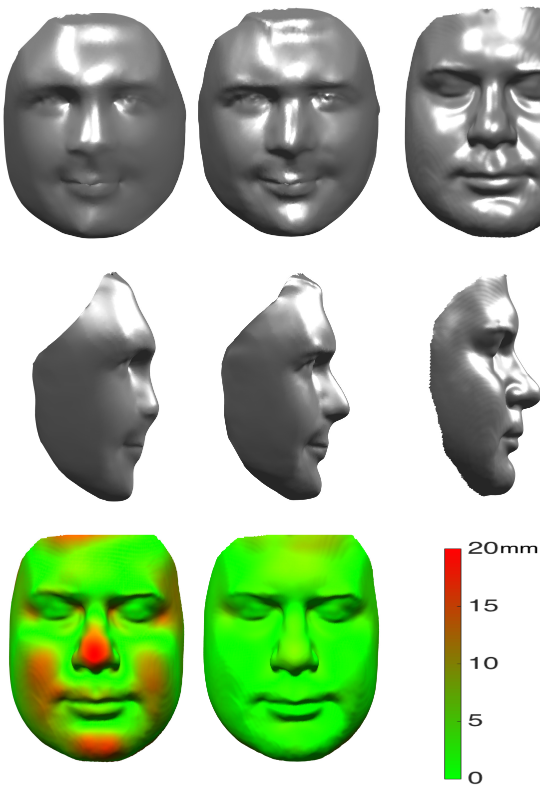

Figure 4: *Distance visualization between a ground truth 3D face acquired using an Artec Eva 3D scanner, and reconstructed faces without (left) and with (right) the silhouette constraints. Red indicates larger distances and green smaller ones. Using silhouette constraints leads to smaller errors in particular in the area around the nose. The range of the distance is 0-20mm from green to red. The 1st column are the results from [RTL15], the 2nd column are our results.*

We plot our results in Figure 6. To compare the results , we re-implement the method in [RTL15]. We test their method with the same dataset used in our implementation. We also generated results with the code made available by Roth et al. [RTL15], but unfortunately this produced inferior results as shown in Figure 5 . Since they only provide a binary executable, we could not investigate the cause of this issue further.

Figure 5: *The 2nd column are the results from the code made available from [RTL15], the 3rd column are our re-implementation results.*

6. Conclusions

We described a method for unconstrained 3D face reconstruction for image collections of individuals dowloaded from the internet, which exhibit unknown illumination, pose, and facial expressions. Our key idea is to introduce silhouette constraints, which are iteratively extracted and updated on the reconstructed 3D mesh, and put into correspondence with edges in each input image. We demonstrate that taking into account our silhouette constraints generally leads to more detailed and accurate 3D reconstructions. In the future, we believe that unconstrained face recognition can be further improved by using dynamic expression models, instead of a single template mesh with a neutral facial expression.

7. Acknowledgement

This work was supported by Swiss National Science Foundation 200021_156253.

References

[BWP13] BOUAZIZ S., WANG Y., PAULY M.: Online modeling for realtime facial animation. *ACM Trans. Graph. 32*, 4 (July 2013), 40:1–40:10. URL: http://doi.acm.org/10.1145/2461912.2461976, doi:10.1145/2461912.2461976. 2

[Can86] CANNY J.: A computational approach to edge detection. In *IEEE Transactions on Pattern Analysis and Machine Intelligence* (June 1986), vol. 8, pp. 679–698. 3

[CBZB15] CAO C., BRADLEY D., ZHOU K., BEELER T.: Real-time high-fidelity facial performance capture. *ACM Trans. Graph. 34*, 4 (July 2015), 46:1–46:9. URL: http://doi.acm.org/10.1145/2766943, doi:10.1145/2766943. 2

[CHZ14] CAO C., HOU Q., ZHOU K.: Displaced dynamic expression regression for real-time facial tracking and animation. *ACM Trans. Graph. 33*, 4 (July 2014), 43:1–43:10. URL: http://doi.acm.org/10.1145/2601097.2601204, doi:10.1145/2601097.2601204. 2

[CWLZ13] CAO C., WENG Y., LIN S., ZHOU K.: 3d shape regression for real-time facial animation. *ACM Trans. Graph. 32*, 4 (July 2013), 41:1–41:10. URL: http://doi.acm.org/10.1145/2461912.2462012, doi:10.1145/2461912.2462012. 2

[CWZ*14] CAO C., WENG Y., ZHOU S., TONG Y., ZHOU K.: Facewarehouse: A 3d facial expression database for visual computing. *IEEE Transactions on Visualization and Computer Graphics 20*, 3 (March 2014), 413–425. doi:10.1109/TVCG.2013.249. 2

[GZC*16] GARRIDO P., ZOLLHOEFER M., CASAS D., VALGAERTS L., VARANASI K., PEREZ P., THEOBALT C.: Reconstruction of personalized 3d face rigs from monocular video. 28:1–28:15. 2

[KS14] KAZEMI V., SULLIVAN J.: One millisecond face alignment with an ensemble of regression trees. In *CVPR* (2014). 4

[KSS11] KEMELMACHER-SHLIZERMAN I., SEITZ S. M.: Face reconstruction in the wild. In *Proceedings of the 2011 International Conference on Computer Vision* (Washington, DC, USA, 2011), ICCV '11, IEEE Computer Society, pp. 1746–1753. URL: http://dx.doi.org/10.1109/ICCV.2011.6126439, doi:10.1109/ICCV.2011.6126439. 1, 2

[LC05] LIU X., CHEN T.: Pose-robust face recognition using geometry assisted probabilistic modeling. In *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)* (June 2005), vol. 1, pp. 502–509 vol. 1. doi:10.1109/CVPR.2005.276. 1

[LYYB13] LI H., YU J., YE Y., BREGLER C.: Realtime facial animation with on-the-fly correctives. *ACM Trans. Graph. 32*, 4 (July 2013), 42:1–42:10. URL: http://doi.acm.org/10.1145/2461912.2462019, doi:10.1145/2461912.2462019. 1, 2

[LZS04] L. ZHANG N. SNAVELY B. C., SEITZ S. M.: Spacetime faces: high resolution capture for modeling and animation. *ACM Trans. Graph 23* (Aug. 2004), 548–558. 4

[MDSB03] MEYER M., DESBRUN M., SCHRÖDER P., BARR A. H.: *Discrete Differential-Geometry Operators for Triangulated 2-Manifolds*. Springer Berlin Heidelberg, Berlin, Heidelberg, 2003, pp. 35–57. URL: http://dx.doi.org/10.1007/978-3-662-05105-4_2, doi:10.1007/978-3-662-05105-4_2. 2, 3

[RTL15] ROTH J., TONG Y., LIU X.: Unconstrained 3d face reconstruction. In *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (June 2015), pp. 2606–2615. doi:10.1109/CVPR.2015.7298876. 1, 2, 3, 4, 5, 7

[SCOL*04] SORKINE O., COHEN-OR D., LIPMAN Y., ALEXA M., RÖSSL C., SEIDEL H.-P.: Laplacian surface editing. In *Proceedings of the 2004 Eurographics/ACM SIGGRAPH Symposium on Geometry Processing* (New York, NY, USA, 2004), SGP '04, ACM, pp. 175–184. URL: http://doi.acm.org/10.1145/1057432.1057456, doi:10.1145/1057432.1057456. 2

[SKSS14] SUWAJANAKORN S., KEMELMACHER-SHLIZERMAN I., SEITZ S. M.: *Total Moving Face Reconstruction*. Springer International Publishing, Cham, 2014, pp. 796–812. URL: http://dx.doi.org/10.1007/978-3-319-10593-2_52, doi:10.1007/978-3-319-10593-2_52. 2

[TZS*16] THIES J., ZOLLHÖFER M., STAMMINGER M., THEOBALT C., NIESSNER M.: Face2face: Real-time face capture and reenactment of rgb videos. In *Proc. Computer Vision and Pattern Recognition (CVPR), IEEE* (2016). 2

[WBLP11] WEISE T., BOUAZIZ S., LI H., PAULY M.: Realtime performance-based facial animation. *ACM Trans. Graph. 30*, 4 (July 2011), 77:1–77:10. URL: http://doi.acm.org/10.1145/2010324.1964972, doi:10.1145/2010324.1964972. 2

[WYWS06] WANG J., YIN L., WEI X., SUN Y.: 3d facial expression recognition based on primitive surface feature distribution. In *2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'06)* (2006), vol. 2, pp. 1399–1406. doi:10.1109/CVPR.2006.14. 1

[YLYL13a] YAN J., LEI Z., YI D., LI S. Z.: Learn to combine multiple hypotheses for accurate face alignment. In *Computer Vision Workshops (ICCVW), 2013 IEEE International Conference on* (Dec 2013), pp. 392–396. doi:10.1109/ICCVW.2013.126. 2

[YLYL13b] YAN J., LEI Z., YI D., LI S. Z.: Learn to combine multiple hypotheses for accurate face alignment. In *Computer Vision Workshops (ICCVW), 2013 IEEE International Conference on* (Dec 2013), pp. 392–396. doi:10.1109/ICCVW.2013.126. 2

[ZLM09] Z. LIN M. C., MA Y.: The augmented lagrange multiplier method for exact recovery of corrupted low-rand matrices. In *Technical Report UILU-ENG-09-2215* (2009). 2

Figure 6: *Comparison results of different individuals between the state of the art technique by Roth et al. [RTL15] and our approach. The 2nd and 5th column are the results from [RTL15], the 3rd and 6th column are our results. In our results, the nose and chin areas generally fit better to the real image. Our results tend to exhibit less oversmoothing and more geometric detail also around the eyes.*