Head Reconstruction from Internet Photos

Shu Liang, Linda G. Shapiro, Ira Kemelmacher-Shlizerman

Computer Science & Engineering Department,
University of Washington
{liangshu,shapiro,kemelmi}@cs.washington.edu

Abstract. 3D face reconstruction from Internet photos has recently produced exciting results. A person's face, e.g., Tom Hanks, can be modeled and animated in 3D from a completely uncalibrated photo collection. Most methods, however, focus solely on face area and mask out the rest of the head. This paper proposes that head modeling from the Internet is a problem we can solve. We target reconstruction of the rough shape of the head. Our method is to gradually "grow" the head mesh starting from the frontal face and extending to the rest of views using photometric stereo constraints. We call our method boundary-value growing algorithm. Results on photos of celebrities downloaded from the Internet are presented.

Keywords: Internet photo collections, head modeling, in the wild, unconstrained 3D reconstruction, uncalibrated

1 Introduction

"If two heads are better than one, then what about double chins? On that note, I will help myself to seconds." —Jarod Kintz

Methods that reconstruct 3D models of people's heads from images need to account for varying 3D pose, lighting, non-rigid changes due to expressions, relatively smooth surfaces of faces, ears and neck, and finally, the hair. Great reconstructions can be achieved nowadays in case the input photos are captured in a calibrated lab setting or semi-calibrated setup where the person has to participate in the capturing session (see related work). Reconstructing from Internet photos, however, is an open problem due to the high degree of variability across uncalibrated photos; lighting, pose, cameras and resolution change dramatically across photos. In recent years, reconstruction of *faces* from the Internet have received a lot of attention. All face-focused methods, however, mask out the head using a fixed face mask and focus only on the face area. For real-life applications, we must be able to reconstruct a full head.

So what is it there to reconstruct except for the face? At the minimum, to create full head models we need to be able to reconstruct the ears, and at least part of the neck, chin, and overall head shape. Additionally, hair reconstruction is a difficult problem. One approach is to use morphable model methods. These,

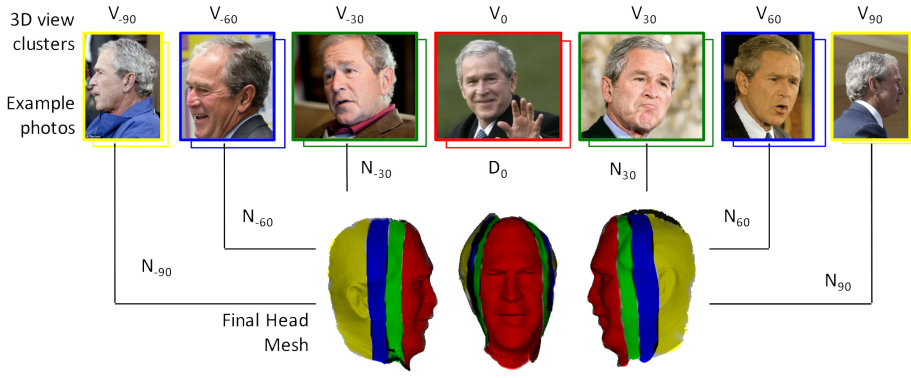

Fig. 1. By looking at the top row photos we can imagine how Bush's head shape looks in 3D; however, existing methods fail to do so on Internet photos, due to such facts as inconsistency of lighting, challenging segmentation,and expression variation. Given many more photos per 3D view (hundreds), however, we show that a rough full head model can be reconstructed. The head mesh is divided into 7 parts, where each part is reconstructed from a different view cluster while being constrained by the neighboring view clusters.

however, do not fit the head explicitly but instead use fitting based on the face and provide a mostly average (non-personalized) bald model for the head.

This paper addresses the new direction of *head* reconstruction directly from Internet data. We propose an algorithm to create a rough head shape, and frame the problem as follows. Given a photo collection, obtained by searching for photos of a specific person on Google image search, we would like to reconstruct a 3D model of that person's head. Just like [1] (that focused only on the face area) we aim to reconstruct an average rigid model of the person from the whole collection. This model can be then used as a template for dynamic reconstruction, e.g., [2] , and hair growing techniques, e.g., [3]. Availability of a template model is essential for those techniques.

Consider the top row photos in Fig. 1. The 3D shape of the head is clearly outlined in the different views (different 3D poses). However, if we are given only one or two photos per view, the problem is still very challenging due to lighting inconsistency across views, difficulty in segmenting the face profile from the background, and challenges in merging the images across views. Our key idea is that with many more (hundreds) of photos per 3D view, the challenges can be overcome. For celebrities, we can easily acquire such collections from the Internet; for others, we can extract such photos from Facebook or from mobile photos.

Our method works as follows: A person's photo collection is divided to clusters of approximately the same azimuth angle of the 3D pose. Given the clusters, a depth map of the frontal face is reconstructed, and the method gradually grows the reconstruction by estimating surface normals per view cluster and then con-

straining using boundary conditions coming from neighboring views. The final result is a head mesh of the person that combines all the views.

2 Related Work

The related work is in calibrated and semi-calibrated setting for head reconstruction, and uncalibrated settings for face reconstruction.

Calibrated head modeling has achieved amazing results over the last decade [4–6]. Calibrated methods require a person to participate in a capturing session to achieve good results. These typically take as input a video with relatively constant lighting, and large pose variation across the video. Examples include non rigid structure from motion methods [7, 8], multiview methods [9, 10], dynamic kinect fusion [11], and RGB-D based methods [12, 13].

Reconstruction of people from Internet photos recently achieved good results; [14] showed that it is possible to reconstruct a face from a single Internet photo using a template model of a different person. [1] later proposed a photometric stereo method to reconstruct a face from many Internet photos of the same individual. [15] showed that photometric stereo can be combined with face landmark constraints, and recent work has shown that 3D dynamic shape [2, 16, 17] and texture [18] can be recovered from Internet photos.

One way to approach the uncalibrated head reconstruction problem is to use the morphable model approach. With morphable models [19, 20], the face is fitted to a linear space of 200 face scans, and the head is reconstructed from the linear space as well. In practice, morphable model methods work well for face tracking [21, 22]. However, there is no actual fitting of the head, ears, and neck of the person to the model, but rather an approximation derived from the face; thus the reconstructed model is not personalized. A morphable model for ears [23] was proposed, but it was not applied to uncalibrated Internet photos.

Hair modeling requires a multiview calibrated setup [24, 25] or can be done from a single photo by fitting to a database of synthetic hairs [3], or by fitting helices [26, 27]. Hair reconstruction methods assume that the user marks hair strokes or that a rough model of the head, ears and face is provided. The goal of this paper is to provide such a rough head shape model; thus our method is complementary to hair modeling techniques.

3 Overview

We denote the set of photos in a view cluster as V_i. Photos in the same view cluster have approximately the same 3D pose and azimuth angle. Specifically, we divided the photos into 7 clusters with azimuths: $i = 0, -30, 30, -60, 60, -90, 90$. Figure 2 shows the averages of each cluster after rigid alignment using fiducial points (1st row) and after subsequent alignment using the Collection Flow method [28] (2nd row), which calculates optical flow for each cluster photo to the cluster average. A key observation is that each view cluster has one particularly well-reconstructed head area, e.g., the ears in views 90 and -90 are sharp

while blurry in other views. Since our goal is to create a full head mesh, our algorithm will combine the optimal parts from each view into a single model. This is illustrated in Figure 1.

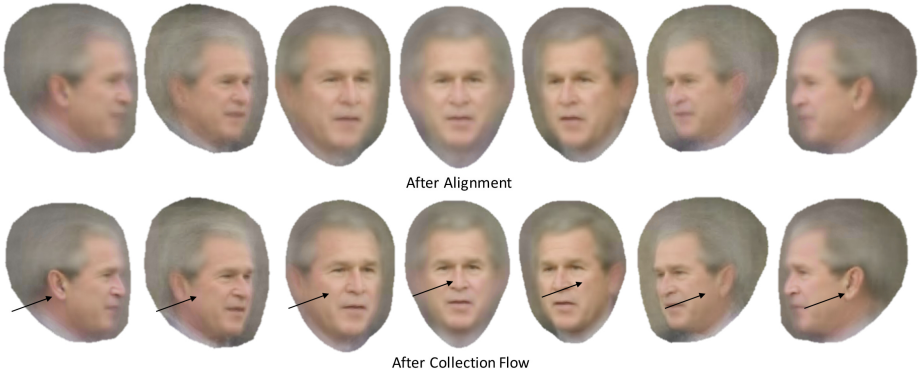

Fig. 2. Averages of view clusters' photos after rigid alignment (1st row) and after collection flow (2nd row). The arrows visualize head parts that are sharper in each view, e.g., the ear is sharpest in 90 and -90 degrees (left and right). The key idea is to use the sharp (well-aligned) parts from the corresponding views to create an optimal mesh reconstruction.

It was shown in previous work that the face can be reconstructed from frontal photos using Photometric Stereo [1]. Thus, one way to implement our idea, of combining views into a single mesh, would be to reconstruct shape from each view cluster independently and then stitch them together. This turned out to be challenging as the individual shapes are reconstructed up to linear ambiguities. Although the photos are divided into pose clusters, the precise pose for each pose cluster is unknown. For example, V_{30} could have a variance from 25 to 35 in the azimuth rotation angle, depending on the dominant pose of the image cluster. This misalignment will also increase the difficulty of stitching all the views. We solve those challenges by growing the shape in stages works well. We begin by describing estimation of surface normals and a depth map for view cluster V_0 (frontal view) in section 4. This will be the initialization for our algorithm. In section 5, we describe how each view cluster uses its own photos and the depth of its neighbors to contribute to the creation of a full head mesh. Data acquisition and alignment details are given in the experiments section (Section 6).

4 Head Mesh Initialization

Our goal is to reconstruct the head mesh M. We begin by estimating a depth map and surface normals of the frontal cluster V_0, and assign each reconstructed

pixel to a vertex of the mesh. The depth map is estimated by extending the method of [1] to capture more of the head in the frontal face photos, i.e. , we extend the reconstruction mask to a bigger area to capture the chin, part of the neck and some of the hair. The algorithm is as follows:

1. **Dense 2D alignment:** Photos are first rigidly aligned using 2D fiducial points as the pipeline of [29]. The head region including neck and shoulder in each image is segmented using semantic segmentation by [30]. Then Collection Flow [28] is run on all the photos in V_0 to densely align them to the average photo of that set. Note that the segementation works remarkably well on most photos. The challenging photos do not affect our method; given that the majority of the photos are segmented well, Collection Flow will correct for inconsistencies. Also, Collection Flow helps overcome differences in hair style by warping all the photos to the dominant style. See more details about alignment in Section 6.

2. **Surface normals estimation:** We used a template face mask to find the face region on all the photos. Photometric Stereo (PS) is then applied to the face region of the flow-aligned photos. The face region of the photos are arranged in an $n \times p_k$ matrix Q, where n is the number of images and p_k is the number of face pixels determined by the template facial mask. Rank-4 PCA is computed to factorize into lighting and normals: $Q = LN$. After we get the lighting estimation L for each photo, we can compute N for all p head pixels including ear, chin and hair regions.

   Two key components that made PS work on uncalibrated head photos are:

   1) resolving the Generalized Bas-Relief (GBR) ambiguity using a template 3D face of a different individual, i.e., $\min_A ||N_{\text{template}} - AN_{face}||^2$,

   2) using a per-pixel surface normal estimation, where each point uses a different subset of photos to estimate the normal. We follow the per-pixel surface estimation idea as in previous work, i.e., given the initial lighting estimate L, the normal is computed per point by selecting a subset of Q's rows that satisfy the re-projection constraint. In the full head case, we extend it to handle cases when the head is partially cropped out, by adding a constraint that a photo participates in normal estimation if it satisfies both the reprojection constraint and is inside the desired head area, i.e., part of the segmentation result from [30]. If the number of selected subset images is not enough (less than $n/3$), we will not use them in our depth map estimation step.

3. **Depth map estimation:** The surface normals are integrated to create a depth map D_0 by solving a linear system of equations that satisfy gradient constrains $dz/dx = -n_x/n_y$ and $dz/dy = -n_x/n_y$ where (n_x, n_y, n_z) are components of the surface normal of each point [31]. Combining these

constraints, for the z-value on the depth map, we have:

$$n_z(z_{x+1,y} - z_{x,y}) = n_x \tag{1}$$

$$n_z(z_{x,y+1} - z_{x,y}) = n_y \tag{2}$$

In the case of $n_z \approx 0$, we use a different constraint,

$$n_y(z_{x,y} - z_{x+1,y}) = n_x(z_{x,y} - z_{x,y+1}) \tag{3}$$

This generate a sparse matrix of $2p \times 2p$ matrix M, and we can solve for:

$$\arg\min_z ||Mz - v||^2 \tag{4}$$

We do a least squares fit to solve for the z-value for each pixel.

Potentially, we could run the same algorithm for each view cluster. This, however, does not perform well, as we will see in the experiments section. Instead we are going to introduce two constraints, which we describe in the next section.

5 Boundary-Value Growing

In this section we describe our "growing" algorithm to complete the side views of the mesh. Starting from the frontal view mesh V_0, we gradually complete more regions of the head in the order of V_{30}, V_{60}, V_{90} and V_{-30}, V_{-60}, V_{-90}. For each view cluster we repeat the same algorithm as in Section 4 with two additional key constraints:

1. **Ambiguity recovery:** Rather than recovering the ambiguity A that arises from $Q = LA^{-1}AN$ using the template model, we use the already computed neighboring cluster, i.e., for $V_{\pm 30}$, N_0 is used, for $V_{\pm 60}$ we use $N_{\pm 30}$, and for $V_{\pm 90}$ we use $N_{\pm 60}$. Specifically, we estimate the out-of-plane pose from our 3D initial mesh V_0 to the average image of pose cluster V_{30} using the method proposed in [2]. We render the rotated mesh V_0' as a reference depth map D_0' to pose cluster V_{30}, accounting for visibility and occlusion using zbuffer. The normals on each projected pixels of D_0' will serve as the reference normals to solve for the GBR ambiguity of the overlapping head region as well as the newly grown head region.

2. **Depth constraint:** In addition to the gradient constraints that are specified in Sec. 4, we modify the boundary constraints from Neumann to Dirichlet. Let Ω_0 be the boundary of D_0'. Then we impose that the part of Ω_0 that intersects the mask of D_{30} will have the same depth values: $D_{30}(\Omega_0) = D_0'(\Omega_0)$. With both boundary constraints and gradient constraints, our optimization function can be written as:

$$\arg\min_z ||Mz - v||^2 + ||Wz - Wz_0||^2 \tag{5}$$

where z_0 is the depth constraint from D'_0, and W is a blend mask with values decreasing from 1 to 0 on the boundary of D'_0. We will get the new vertex positions for grown regions and we can also update vertices on the boundary of the already computed depth map, eliminating the distortion caused by lack of photos and inaccurate n_z. This process is repeated for every neighboring pair of depths.

After each depth stage reconstruction (0,30,60,.. degrees), the estimated depth is projected to the head mesh. By this process, the head is gradually filled in by gathering vertices from all the views.

6 Experiments

We describe the data collection process, alignment, evaluations and comparisons with other methods.

6.1 Data Collection and Processing

We collected around 1,000 photos per person (George Bush, Vladimir Putin, Barack Obama and Hillary Clinton) by searching for photos on Google image search. The numbers of images in each pose cluster are shown in Table 1. We noticed that the numbers of side view photos are usually much smaller than frontal view photos. In order to get more photos, we searched for "Bush shakes hands", "Bush shaking hand", "Bush portrait", "Bush meets" etc. to collect more non-frontal photos. The number of photos in each cluster will affect the final result; we will demonstrate the reconstruction quality vs. number of photos later in this section.

Table 1. Number of photos we used in each pose cluster

Pose	-90	-60	-30	0	30	60	90
Bush	185	62	118	371	113	80	191
Putin	131	58	151	413	121	61	151
Obama	65	51	126	284	177	55	75
Clinton	115	47	114	332	109	61	66

We ran face detection and fiducial detection using IntraFace[32]. For extreme side views, none of the state of the art fiducial detection algorithms was able to perform, and often times the face was not even detected. We therefore manually annotated each photo with 7 fiducials.

Once photos are aligned we run collection flow [28] on each view cluster. For completeness we review the method. The idea is to estimate a lighting subspace from all the photos in a particular cluster V_i via PCA. Then each photo in the

cluster V_i^j is projected to the subspace to produce photo $\hat{V}_i^j$, which has a similar lighting as V_i^j but an average shape. Optical flow is then estimated between V_0^j and its relighted version $\hat{V}_0^j$. The process is iterated over the whole collection. In the end, all photos are warped to approximately average shape; however, they retain their lighting which makes them amenable for photometric stereo methods.

6.2 Results and Evaluation

Fig. 3 shows the reconstruction per view that was later combined to a single mesh. For example, the ear in 90 and -90 views is reconstructed well, while the other views are not able to reconstruct the ear.

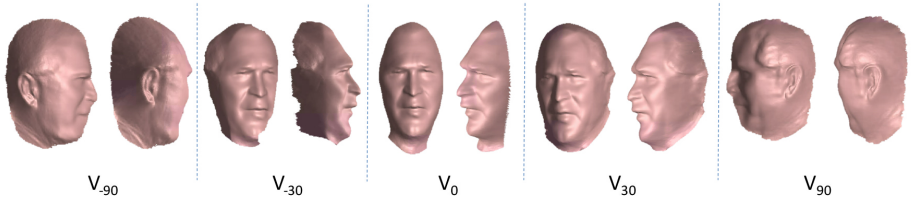

Fig. 3. Individual reconstructions per view cluster, with depth and ambiguity constraints. We can see that the individual views provide different shape components. For each view we show the mesh in two poses.

In Figure 4, we shows how our two key constraints work well in the degree 90 view reconstruction result. Without the correct reference normals and depth constraint, the reconstructed shape is flat and the profile facial region is blurred, which increased the difficulty of aligning it back to the frontal view.

Fig. 5 shows the reconstruction result for 4 subjects, each mesh is rotated to five different views. Note that the back and top part of the head are partly missing due to the lack of photos.

To evaluate how the number of photos affects the reconstruction quality, we took 600 photos for George Bush and estimated pose, lighting, texure for each image. We report the L2 intensity difference between the rendered photos and original photos. We tested our reconstruction method with 1/2, 1/4, 1/8 and 1/16 of the photos in each view cluster (see number of photos per cluster in Table 1.) The method did not work in 1/16 case because some view clusters have less than 10 photos and there was not enough lighting variation within the collection for photometric stereo. Generally, we suggest using more than 100 photos for frontal view. The number of photos in side view clusters can be smaller (but larger than 30) because the side view of a human's head is more rigid than the frontal view.

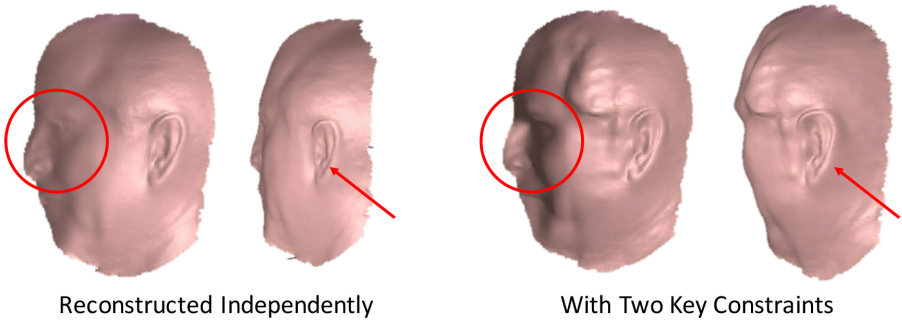

Fig. 4. Comparison between without and with two key constraints. The left two shapes show the two views of 90 degree view shape reconstructed independently without two key constraints. The right two shapes show the two views of our result with two key constraints.

Table 2. Reconstruction Quality vs. Number of Photos

Number of photos	N	N/2	N/4	N/8	N/16
Reprojection Error(intensity)	18.29 ± 4.07	18.70 ± 4.07	18.71 ± 4.07	18.80 ± 4.04	*N/A*

We also rendered a 3D model from the FaceWareHouse dataset [33] with 100 lights and 7 poses. We applied our method on these synthetic photos and got a reconstruction result as shown in Fig 6. Since we use a template 3D model to correct GBR ambiguity, we cannot get the exact scale of the groundtruth. We do not claim that we have recovered the perfect shape, but the result looks reasonable with an average reprojection error of 11.1 ± 5.72.

6.3 Comparison

In Figure 7 we show a comparison to the software FaceGen that implements a morphable model approach. For each person, we manually selected three photos (one frontal view and two side view photos) and used them as the input for FaceGen. The results of FaceGen are too averaged out and not personalized. Note that their ears look the same as each other.

We also tried the Space Carving method [34]. For each subject, we manually selected about 30 photos in different poses with a neutral expression. We used the segmentation result obtained from Section 4 as the silhouette. We assumed the camera focus length to be 100 and estimated the camera extrinisic parameters using a template 3D head model. We smoothed the carved results using [35] and showed the reconstruction in Figure 8. The Space Carving method can produce

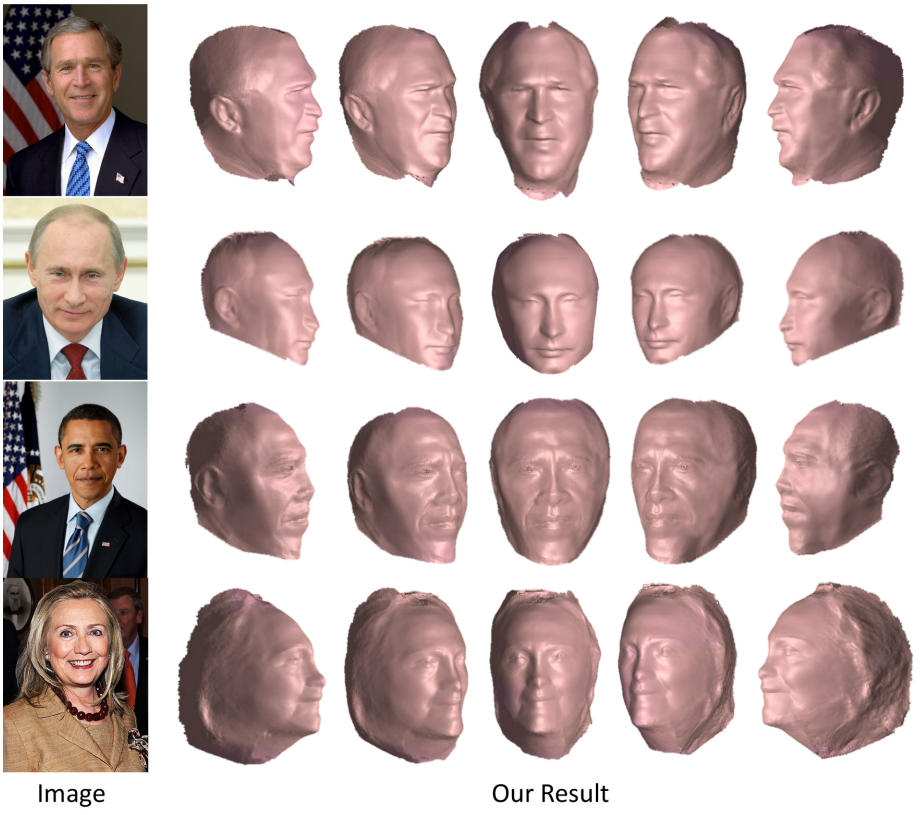

Fig. 5. Final reconstructed mesh rotated to 5 views to show the reconstruction from all sides. Each color image is an example image among our around 1,000 photo collection for each person.

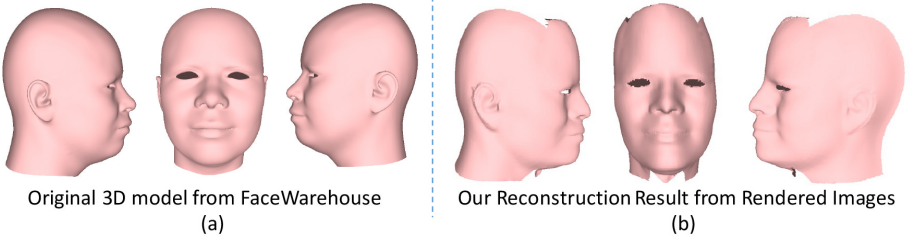

Fig. 6. Reconstruction result from the synthetic photos rendered from a 3D model in FaceWarehouse. The left three shapes are the $-90, 0, 90$ views for the groundtruth shape, and the right three shapes are our reconstruction result.

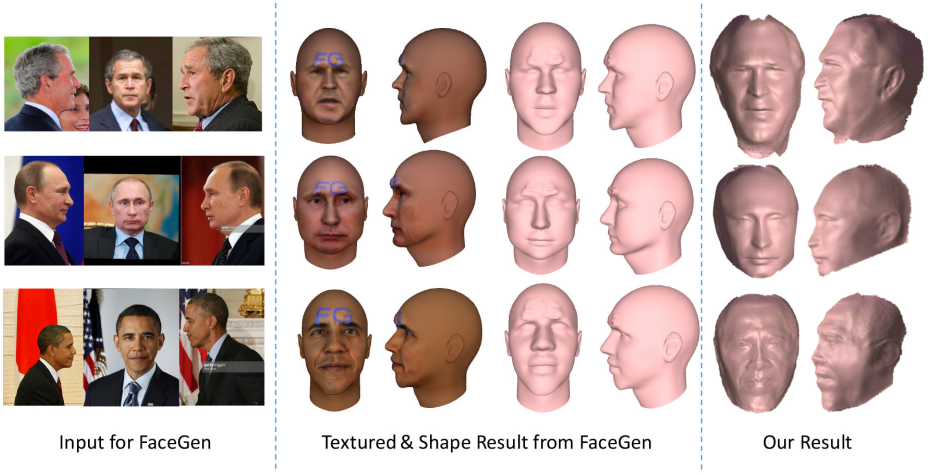

Fig. 7. Comparison to FaceGen (morphable model). We show the textured results and shape results from FaceGen in the middle and our results are on the right as comparisons. Note that the head shape reconstructed by morphable models is average like and not personalized. Additionally, texture hides shape imperfections.

a rough shape of the head. Increasing the number of photos to use does not improve the result.

We have also experimented with VisualSfM [36], but the software could not find enough feature points to run a structure from motion method. This is probably due to the lighting variation and expression change in the photo collection. Similarly, we have tried http://www.123dapp.com/catch, and it was not able to reconstruct from such photos.

Table 3. Reprojection error from 3 reconstruction methods.

Reprojection error	FaceGen	Visual Hull	Our method
Bush	20.1 ± 4.84	19.6 ± 3.55	18.3 ± 4.04
Putin	20.1 ± 4.84	17.2 ± 4.68	15.1 ± 5.06
Obama	21.5 ± 4.62	20.7 ± 4.58	19.7 ± 4.40

For a quantitative comparison, for each person, we calculated the reprojection error of the shapes from three methods (ours, Space Carving and FaceGen) to 600 photos in different poses and lighting variations. The 3D shape comes from each reconstruction method. The albedo all comes from average shapes of our clusters, since the Space Carving method and the FaceGen results do not include albedos. The average reprojection error is shown in Table 3. The error map of

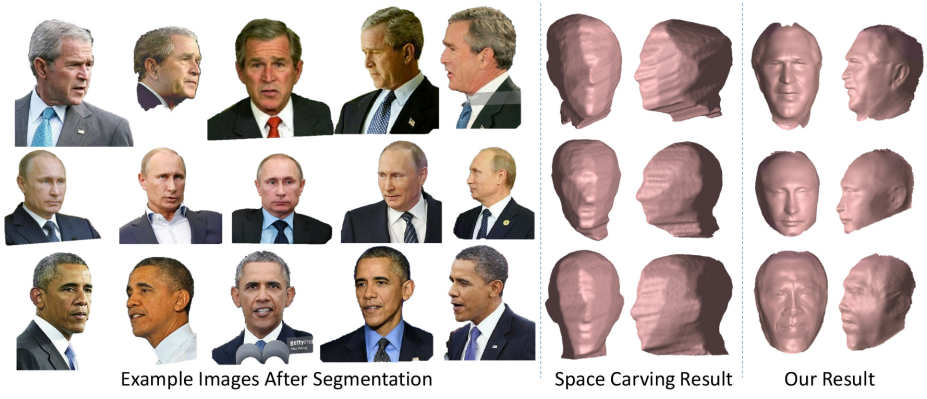

Fig. 8. Comparison to Space Carving method. 5 example segmented images are shown on the left for each person. The segmentations were used as silhouettes. We used around 30 photos per person.

an example image is shown in Fig 9. We calculated the error for the overlapping pixels of the three rendered images. Notice that the shapes from FaceGen and Space Carving might look good from the frontal view, but they are not correct when rotating to the target view. See how different the ear part is in the figure.

In future work, we will test the algorithm on more people. Collecting side view images is time consuming. Currently, there are no sets of Internet photos with their corresponding 3D models, thus it is challenging to evaluate quantitatively. We would like to help to solve that by providing our dataset. Furthermore, our GBR ambiguity is just roughly solved by a template model, so the scale might be not exactly the same as the actual mesh. We do not claim to have recovered the perfect shape, but rather show that it is possible to do so from Internet photos, and to encourage further research.

7 Discussion

We have shown the first results of head reconstructions from Internet photos. Our method has a number of limitations. First, we assume a Lambertian model for surface reflectance. While this works well, accounting for specularities should improve results. Second, fiducials for side views were labeled manually; we hope that this application will encourage researchers to solve the challenge of side view fiducial detection. Third, we have not reconstructed a complete model; the top of the head is missing. To solve this we would need to add photos with different elevation angles, rather than just focusing on the azimuth change.

We see several possible extensions to our method. The two we are most excited about are 1) reconstructing 3D non-rigid motion that includes the head part (not only face, as was done until now), and 2) combining with hair growing

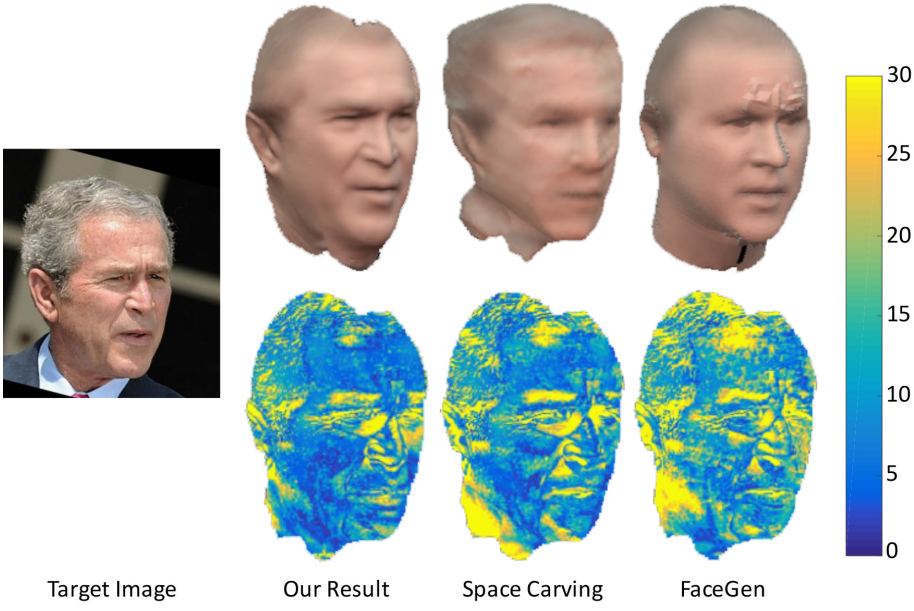

Fig. 9. Visualization of the reprojection error for 3 methods.

methods that can use our reconstructed shape as initialization, e.g., in [26] the template was produced manually.

References

1. Kemelmacher-Shlizerman, I., Seitz, S.M.: Face reconstruction in the wild. In: Computer Vision (ICCV), 2011 IEEE International Conference on, IEEE (2011) 1746–1753
2. Suwajanakorn, S., Kemelmacher-Shlizerman, I., Seitz, S.M.: Total moving face reconstruction. In: Computer Vision–ECCV 2014. (2014)
3. Hu, L., Ma, C., Luo, L., Li, H.: Single-view hair modeling using a hairstyle database. ACM Transactions on Graphics (TOG) **34**(4) (2015) 125
4. Debevec, P.: The light stages and their applications to photoreal digital actors. SIGGRAPH Asia (2012)
5. Alexander, O., Fyffe, G., Busch, J., Yu, X., Ichikari, R., Jones, A., Debevec, P., Jimenez, J., Danvoye, E., Antionazzi, B., et al.: Digital ira: Creating a real-time photoreal digital actor. In: ACM SIGGRAPH 2013 Posters, ACM (2013) 1
6. Beeler, T., Bickel, B., Beardsley, P., Sumner, B., Gross, M.: High-quality single-shot capture of facial geometry. ACM Transactions on Graphics (TOG) **29**(4) (2010) 40
7. Agudo, A., Montiel, J., de Agapito, L., Calvo, B.: Online dense non-rigid 3d shape and camera motion recovery. In: BMVC. (2014)

8. Garg, R., Roussos, A., Agapito, L.: Dense variational reconstruction of non-rigid surfaces from monocular video. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. (2013) 1272–1279
9. Tanskanen, P., Kolev, K., Meier, L., Camposeco, F., Saurer, O., Pollefeys, M.: Live metric 3d reconstruction on mobile phones. In: Proceedings of the IEEE International Conference on Computer Vision. (2013) 65–72
10. Ichim, A.E., Bouaziz, S., Pauly, M.: Dynamic 3d avatar creation from hand-held video input. ACM Transactions on Graphics (TOG) **34**(4) (2015) 45
11. Newcombe, R.A., Fox, D., Seitz, S.M.: Dynamicfusion: Reconstruction and tracking of non-rigid scenes in real-time. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. (2015) 343–352
12. Thies, J., Zollhoefer, M., Niessner, M., Valgaerts, L., Stamminger, M., Theobalt, C.: Real-time expression transfer for facial reenactment. ACM Transactions on Graphics (Proc. SIGGRAPH Asia) (2015)
13. Zollhöfer, M., Nießner, M., Izadi, S., Rehmann, C., Zach, C., Fisher, M., Wu, C., Fitzgibbon, A., Loop, C., Theobalt, C., et al.: Real-time non-rigid reconstruction using an rgb-d camera. ACM Transactions on Graphics (TOG) **33**(4) (2014) 156
14. Kemelmacher-Shlizerman, I., Basri, R.: 3d face reconstruction from a single image using a single reference face shape. Pattern Analysis and Machine Intelligence, IEEE Transactions on **33**(2) (2011) 394–405
15. Roth, J., Tong, Y., Liu, X.: Unconstrained 3d face reconstruction. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. (2015) 2606–2615
16. Garrido, P., Valgaerts, L., Wu, C., Theobalt, C.: Reconstructing detailed dynamic face geometry from monocular video. ACM Trans. Graph. **32**(6) (2013) 158–1
17. Shi, F., Wu, H.T., Tong, X., Chai, J.: Automatic acquisition of high-fidelity facial performances using monocular videos. ACM Transactions on Graphics (TOG) **33**(6) (2014) 222
18. Suwajanakorn, S., Seitz, S.M., Kemelmacher-Shlizerman, I.: What makes tom hanks look like tom hanks. In: Proceedings of the IEEE International Conference on Computer Vision. (2015) 3952–3960
19. Blanz, V., Vetter, T.: A morphable model for the synthesis of 3d faces. In: Proceedings of the 26th annual conference on Computer graphics and interactive techniques, ACM Press/Addison-Wesley Publishing Co. (1999) 187–194
20. Hsieh, P.L., Ma, C., Yu, J., Li, H.: Unconstrained realtime facial performance capture. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. (2015) 1675–1683
21. Shapiro, A., Feng, A., Wang, R., Li, H., Bolas, M., Medioni, G., Suma, E.: Rapid avatar capture and simulation using commodity depth sensors. Computer Animation and Virtual Worlds **25**(3-4) (2014) 201–211
22. Li, H., Yu, J., Ye, Y., Bregler, C.: Realtime facial animation with on-the-fly correctives. ACM Trans. Graph. **32**(4) (2013) 42–1
23. Bustard, J.D., Nixon, M.S.: 3d morphable model construction for robust ear and face recognition. In: Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on, IEEE (2010) 2582–2589
24. Luo, L., Li, H., Rusinkiewicz, S.: Structure-aware hair capture. ACM Transactions on Graphics (TOG) **32**(4) (2013) 76
25. Hu, L., Ma, C., Luo, L., Li, H.: Robust hair capture using simulated examples. ACM Transactions on Graphics (TOG) **33**(4) (2014) 126

26. Chai, M., Luo, L., Sunkavalli, K., Carr, N., Hadap, S., Zhou, K.: High-quality hair modeling from a single portrait photo. ACM Transactions on Graphics (TOG) **34**(6) (2015) 204
27. Chai, M., Wang, L., Weng, Y., Yu, Y., Guo, B., Zhou, K.: Single-view hair modeling for portrait manipulation. ACM Transactions on Graphics (TOG) **31**(4) (2012) 116
28. Kemelmacher-Shlizerman, I., Seitz, S.M.: Collection flow. In: Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on, IEEE (2012) 1792–1799
29. Kemelmacher-Shlizerman, I., Shechtman, E., Garg, R., Seitz, S.M.: Exploring photobios. In: ACM Transactions on Graphics (TOG). Volume 30., ACM (2011) 61
30. Zheng, S., Jayasumana, S., Romera-Paredes, B., Vineet, V., Su, Z., Du, D., Huang, C., Torr, P.H.: Conditional random fields as recurrent neural networks. In: Proceedings of the IEEE International Conference on Computer Vision. (2015) 1529–1537
31. Basri, R., Jacobs, D., Kemelmacher, I.: Photometric stereo with general, unknown lighting. International Journal of Computer Vision **72**(3) (2007) 239–257
32. Xiong, X., Torre, F.: Supervised descent method and its applications to face alignment. In: Proceedings of the IEEE conference on computer vision and pattern recognition. (2013) 532–539
33. Cao, C., Weng, Y., Zhou, S., Tong, Y., Zhou, K.: Facewarehouse: a 3d facial expression database for visual computing. Visualization and Computer Graphics, IEEE Transactions on **20**(3) (2014) 413–425
34. Kutulakos, K.N., Seitz, S.M.: A theory of shape by space carving. International Journal of Computer Vision **38**(3) (2000) 199–218
35. Desbrun, M., Meyer, M., Schröder, P., Barr, A.H.: Implicit fairing of irregular meshes using diffusion and curvature flow. In: Proceedings of the 26th annual conference on Computer graphics and interactive techniques, ACM Press/Addison-Wesley Publishing Co. (1999) 317–324
36. Wu, C.: Visualsfm: A visual structure from motion system. (2011)