UNIVERSITY of WASHINGTON

ECCV 2016

Head Reconstruction from Internet Photos

Shu Liang, Linda G. Shapiro, Ira Kemelmacher-Shlizerman
University of Washington, Computer Science & Engineering

Input: Internet Photos

Output: 3D Model of the Head

Key contribution: head modeling from uncalibrated data is possible.

Results

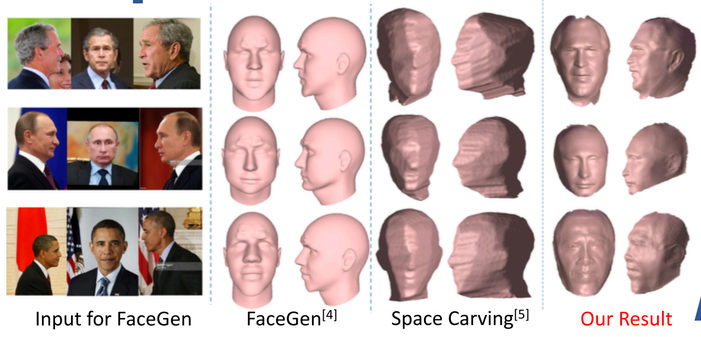
Pose	-90	-60	-30	0	30	60	90
Bush	185	62	118	371	113	80	191
Putin	131	58	151	413	121	61	151
Obama	65	51	126	284	177	55	75
Clinton	115	47	114	332	109	61	66

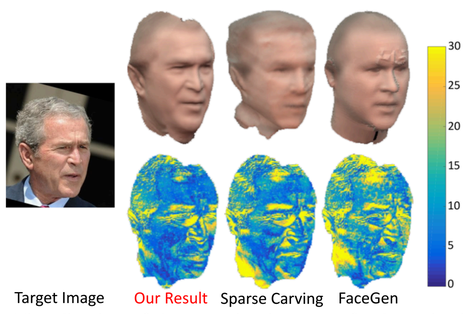
Number of Photos we used in each pose cluster

Number of photos	N	N/2	N/4	N/8	N/16
Reprojection error	18.29±4.07	18.70±4.07	18.71±4.07	18.80±4.04	N/A

Reconstruction Quality vs. Number of Photos

Comparison





Visualization of the reprojection error for 3 methods.

Reprojection error	FaceGen	Space Carving	Our method
Bush	20.6±3.80	19.6±3.55	18.3±4.04
Putin	20.1±4.	17.2±4.68	15.1±5.06
Obama	21.5±4.62	20.7±4.58	19.7±4.40

Reprojection errors for 3 methods. (averaged over 600 photos)

Limitations

We have shown the first results of head reconstructions from Internet photos, but:
1) Lambertian model doesn't capture hair well. We also haven't worked on reconstructing the details. This model could be combined with template based models.
2) Fiducials for side views were labeled manually.
3) 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.

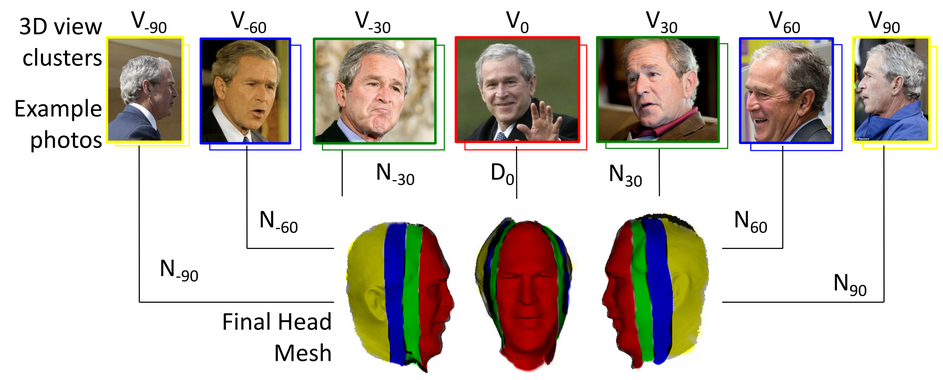
Acknowledge

We thank Samsung, Google, and NSF/Intel grant #1538613 for funding this research.

References

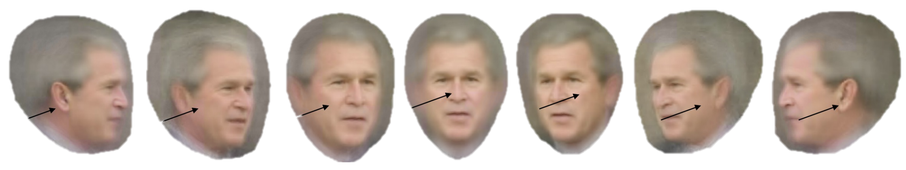
[1] Zheng, S., Jayasumana, S., Romera-Paredes, B., Vineet, V., Su, Z., Du, D., Huang, C., Torr, P.H.: Conditional random fields as recurrent neural networks. In: Proceedings of the IEEE International Conference on Computer Vision. (2015) 1529–1537
[2] Kemelmacher-Shlizerman, I., Seitz, S.M.: Collection flow. In: Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on, IEEE (2012) 1792-1799
[3] Basri, R., Jacobs, D., Kemelmacher, I.: Photometric stereo with general, unknown lighting. International Journal of Computer Vision 72(3) (2007) 239-257
[4] FaceGen: http://facegen.com/
[5] Kutulakos, K.N., Seitz, S.M.: A theory of shape by space carving. International Journal of Computer Vision 38(3) (2000) 199–218

1. Pose-cluster-based reconstruction

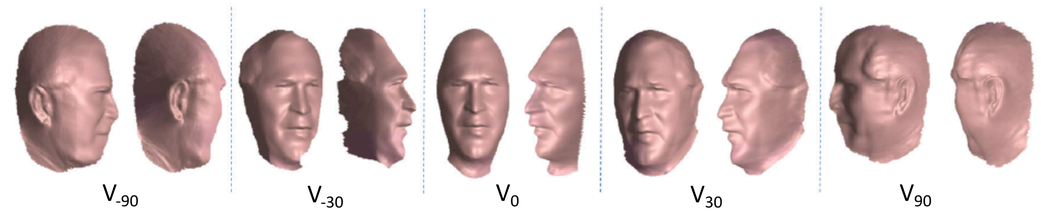


RNN Semantic segmentation[1]

Average Photos after Rigid Alignment



Average Photos after Collection Flow[2]



Observation: each view cluster has one particularly **well**-reconstructed area

2. Photometric Stereo[3] on frontal pose cluster (D_0)

Frontal view photos to n x p matrix Q.

n: number of photos, p: number of pixels of the facial mask. Rank-4 PCA is computed to factorize into lighting and normals: $Q = LN$, with ambiguity, $Q=LA^{-1}AN$.

Resolve the Generalized Bas-Relief (GBR) ambiguity using a template 3D face of a different individual, i.e.,

$$\min_A ||N_{template} - AN_{face}||^2$$

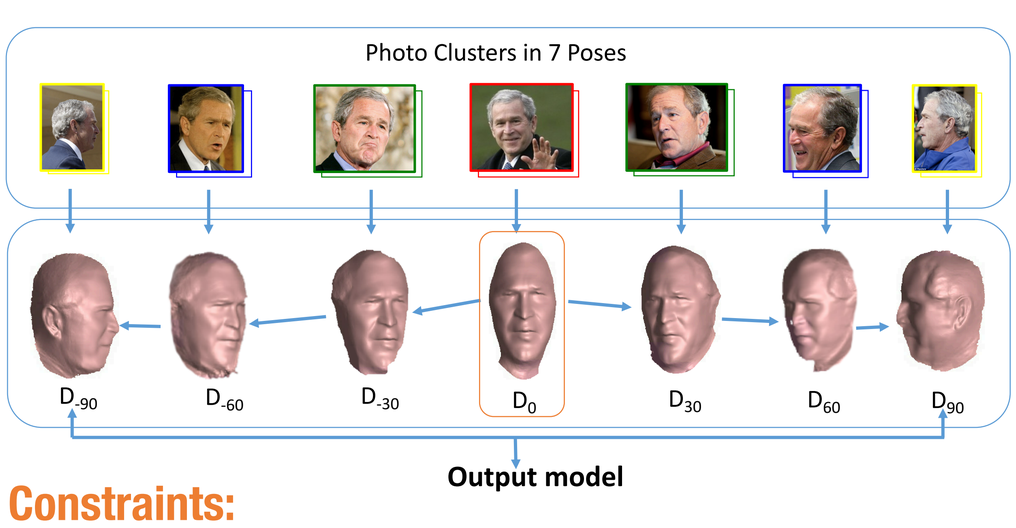
The surface normals are integrated to create D_0 by solving linear equations that satisfy gradient constraints:

$$n_z(D_{x+1,y} - D_{x,y}) = n_x$$
$$n_z(D_{x,y+1} - D_{x,y}) = n_y$$
$$n_y(D_{x,y} - D_{x+1,y}) = n_x(D_{x,y} - D_{x,y+1})$$

This generates a sparse 2p x 2p matrix M, and we can solve for:

$$\arg\min_D ||MD - v||^2$$

3. Boundary-Value Growing



Constraints:

$\arg\min_{A_i} ||N_i - A_i N_j||^2$ GBR ambiguity solved by a neighboring pose.

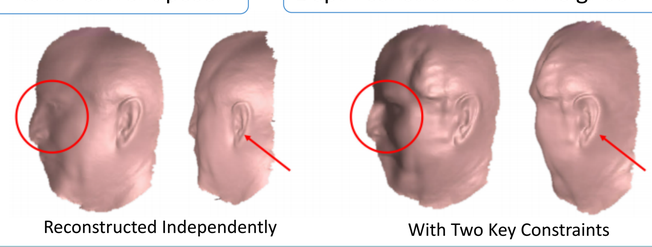
blend mask — depth from a neighboring pose

$\arg\min_{z_i} ||M_i D_i - v_i||^2 + ||W_i D_i - W_i D_j||^2$

e.g. W_{30}

Normal constraints for current pose.

Depth constraints from neighboring pose.



Reconstructed Independently — With Two Key Constraints

dataset is released on our webpage: http://grail.cs.washington.edu/projects/headfromphotos/