3D Face Hallucination from a Single Depth Camera

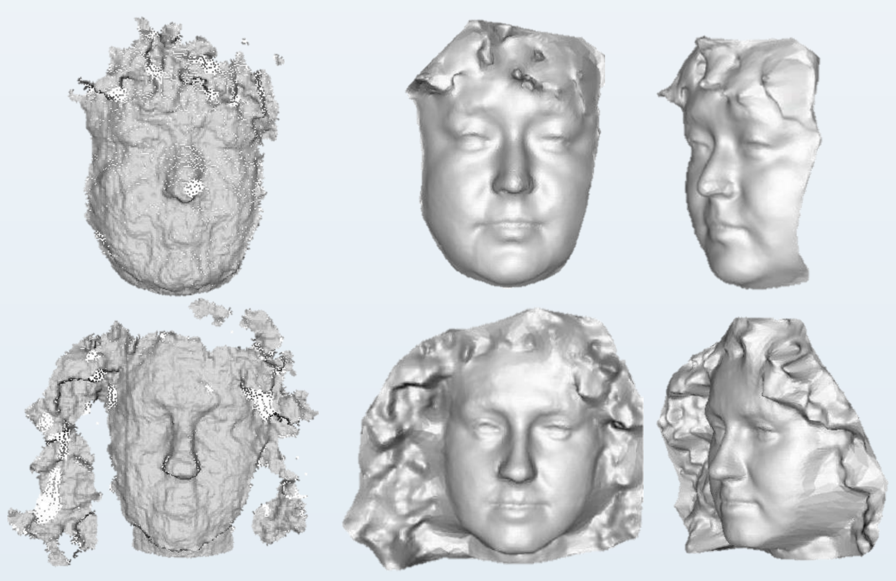
Shu Liang, Ira Kemelmacher-Shlizerman, Linda G. Shapiro

Department of Computer Science& Engineering, University of Washington, Seattle, WA 98195, USA.

Contribution

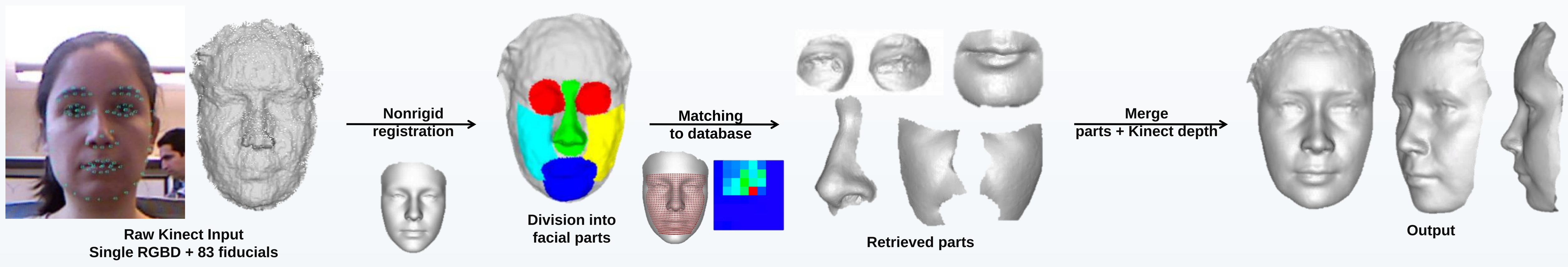
We present an algorithm that takes a single frame of a person's face from a depth camera, e.g., Kinect, and produces a high-resolution 3D mesh of the input face.

- **A single depth frame encodes information about a person's facial features.**
- **Hallucinate details from a high-resolution dataset of 3D face scans.**



Single depth frame Reconstructed 3D mesh

Method



We first align the input RGBD frame to the generic mesh G . Then the input depth is divided into five facial parts via the alignment, and each facial part is matched independently to the dataset resulting in five high-resolution meshes. Finally, the matched meshes are combined with the input into a single mesh to produce the output.

Aligning a Single Depth Frame

- **83 fiducial points on RGB using Face++[1]**
- **Rigid pose alignment via Procrustes analysis[2]**
- **Non-rigid registration[3]**
- **Define five facial parts**

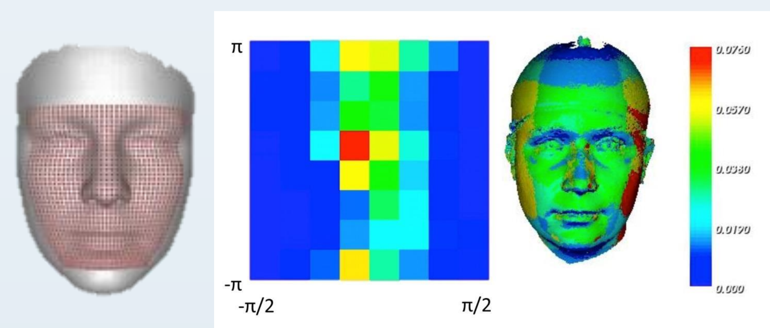
Part-based Matching

Distance function: a combination of pseudo-landmarks and a histogram of azimuth-elevation angles of normals

$$D_{pts}^{j} = \sum_{i=1}^{(m+2)n} \left\| P_i^j - P_i^{input} \right\|^2$$

$$D_{normals} = \chi^2(H^j, H^{input})$$

$$D = D_{pts} + \alpha D_{normals}$$



Merging

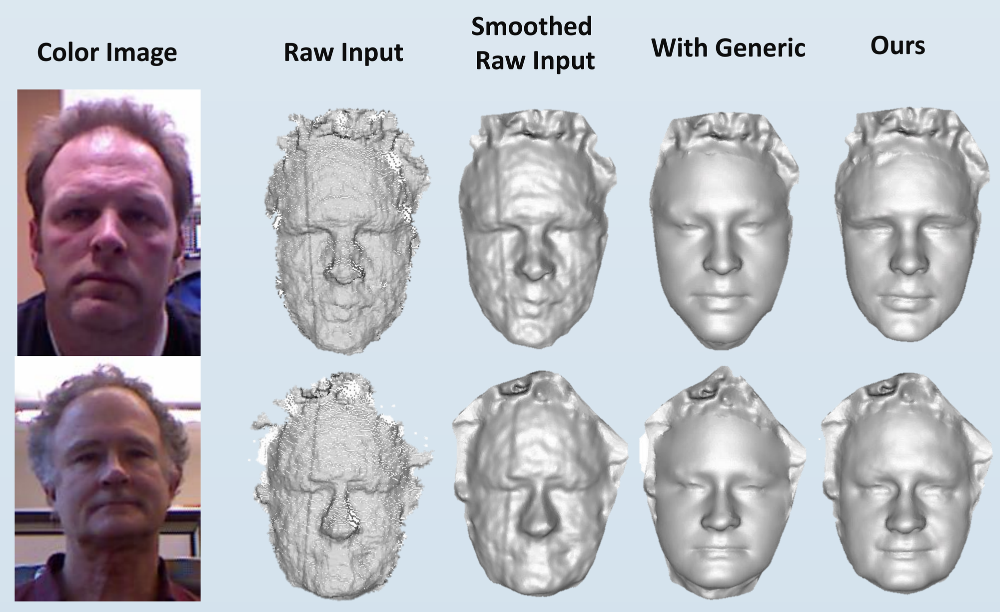
- **Skin region: vertex normal transferred from matched shapes. Hair region: original normal kept**
- **Combine depth and normal[4]**

Dataset

- **A large dataset of high-resolution 3D face meshes in a neutral expression (no texture).**
- **1204 Caucasians, 652 females and 552 males, ages 3 to 40, captured in a neutral expression.**
- **All the meshes in the dataset have been put into dense correspondence using [3].**
- **Each includes 15k-20k vertices.**

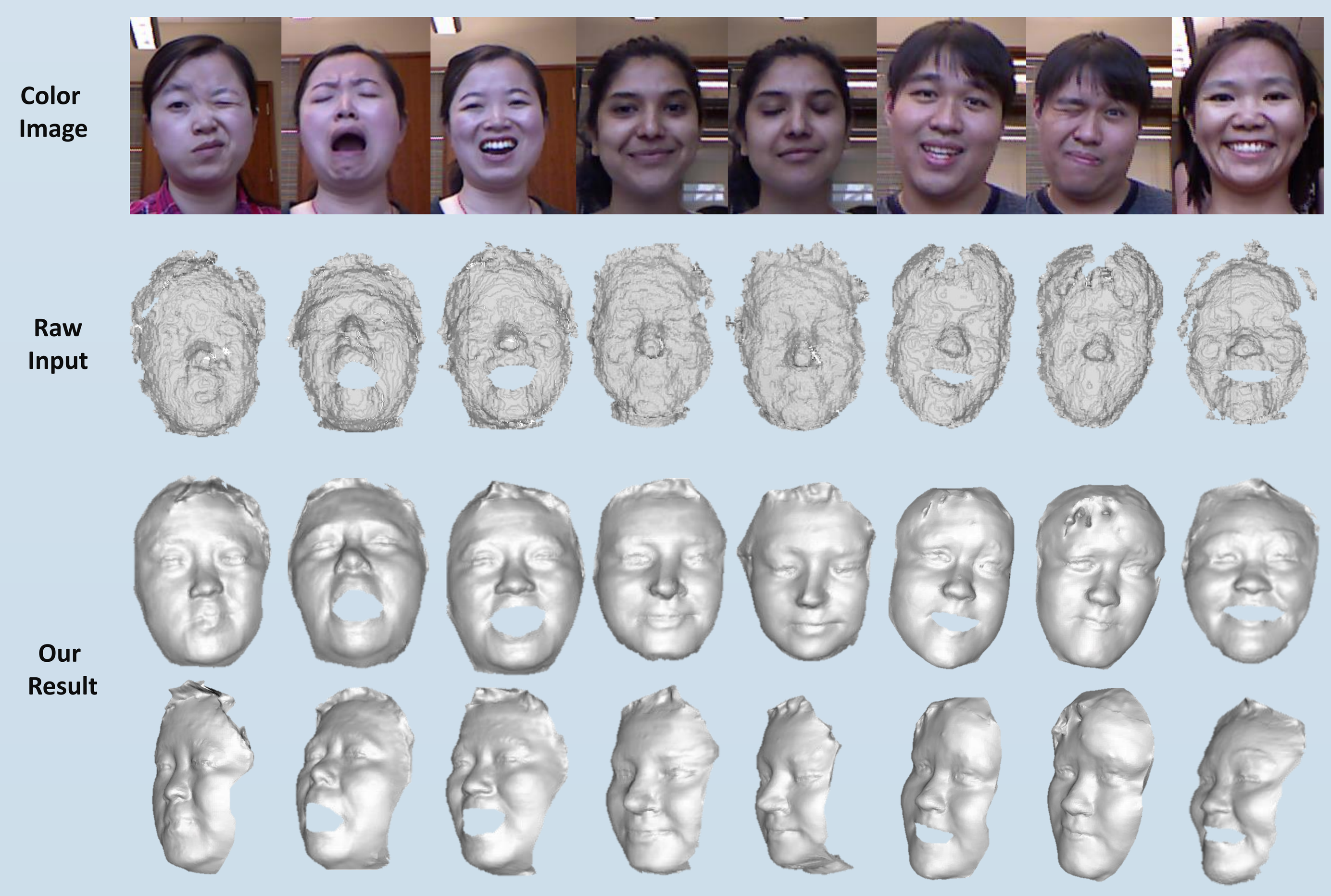
Examples of high-resolution face meshes

Results



Neutral results, compared to details from a generic shape

We used a Microsoft Kinect to capture the inputs in resolution 640 x 480; the face part of the frame was about 100 x 100.



Expression Results

Conclusion

We presented our approach for reconstruction of a high-quality 3D face mesh from a rough, noisy, low-resolution single Kinect depth frame.

Our key contribution is to show that extremely simple part-based matching to a large set of faces enables the creation of remarkably accurate high-resolution meshes of novel people from noisy single-frame input. The resultant meshes can be further used for facial expression modeling, as we also demonstrated.

References:

[1] M. Inc. Face++ research toolkit. www.faceplusplus.com, Dec. 2013.
[2] J. C. Gower. Generalized procrustes analysis. Psychometrika, 40(1):33-51,1975.
[3] B. Allen, B. Curless, and Z. Popovi´c. The space of human body shapes: reconstruction and parameterization from range scans. In ACM Transactions on Graphics (TOG), volume 22, pages 587–594. ACM, 2003
[4] D. Nehab, S. Rusinkiewicz, J. Davis, and R. Ramamoorthi. Efficiently combining positions and normals for precise 3d geometry. In ACM Transactions on Graphics (TOG), volume 24, pages 536–543. ACM, 2005.