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.

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

$$D_{pseudo} = \sum_{i=1}^{M} \left| \theta_{i} - \theta_{0}\right|^2$$

$$D_{nora} = \sum_{j=1}^{N} \left( \mathbf{v}_{j}^{G} - \mathbf{v}_{j}^{G_{0}} \right)^{T} \mathbf{R} \left( \mathbf{v}_{j}^{G} - \mathbf{v}_{j}^{G_{0}} \right)$$

### Merging

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

### Results

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.

### 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.

### 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.

### References: