Towards Precise Completion of Deformable Shapes Supplementary Material

Oshri Halimi*[1], Ido Imanuel*[1], Or Litany[2], Giovanni Trappolini[3], Emanuele Rodolà[3], Leonidas Guibas[2], and Ron Kimmel[1]

[1] Technion - Israel Institute of Technology
[2] Stanford University
[3] Sapienza University of Rome

In this supplementary we provide:

1. Further analysis of our network design
   - An ablation experiment demonstrating the improved reconstruction over a fixed template due to the full shape Q in Section 1.1.
   - Robustness analysis of our trained network in Section 1.2.
   - An experiment exploring the network operation under different types of projections, resulting from elevated viewpoints, in Section 1.3
2. Additional visualizations of the network reconstructions in Section 2.
3. Visualizations of the dense correspondence results from the partial shape to the full shape in Section 3.

1 Analysis

1.1 Comparison with a fixed template baseline

As described in the main manuscript, in order to predict the completion of a partial shape P, our method requires a full reference shape Q of the same subject in an arbitrary pose. We motivate this setting by a requirement for a completion that is faithful to the subject shape. This is different from previous completion methods which can only approximate or hallucinate missing details. Here we support this claim experimentally, by comparing with a baseline that uses a fixed template. Specifically, instead of providing a full shape Q of the same subject as the partial shape P, we provide a *fixed* template T for all inputs. With this modification, the ablation network is trained with the triplets $\{(P_n, T, R_n)\}_{n=1}^{N}$, where N is the size of the training set. At inference time, we use the same template T to make a prediction for a given input part P. We chose the template to be the first subject from the FAUST Projections dataset, in its null pose as shown in Figure 1. Both the original and the fixed-template networks were trained on the FAUST Projections training set, with identical parameters and for the same number of epochs, as described in Section 3.6 in the paper. Table 1 summarizes the prediction errors of both methods, Figure 3 compares the partial correspondence results and Figure 2 shows visual comparison. The results clearly

* equal contribution

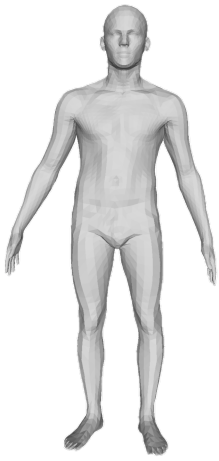

Fig. 1. Constant template used in ablation fixed-template experiment

show the benefit of utilizing the shared geometry between the part and a full non-rigid observation of it. In particular, we receive a noticeable improvement in correspondence prediction as well as a lower reconstruction error across all metrics. Perhaps more importantly, Figure 2 demonstrates the main motivation of our framework: a completion that respects the fine details of the underlying shape. To further emphasize this effect, we magnify the face regions of each shape, showing the loss in intricate details achieved with the alternative training method.

Figure 2 implies how powerful our method is when it comes to the reconstruction of fine details, such as the facial structure and delicate body features. We verify that acquiring access to a full observation in inference time can significantly improve the reliability of the reconstruction for a network trained to utilize such information. In the absence of this full observation at inference time, the ablation network can only utilize the input part and the acquired statistics of the training examples, encoded in the network weights. While this later information can be used for coarse completion, we evidence it is not sufficient for *precise* completion.

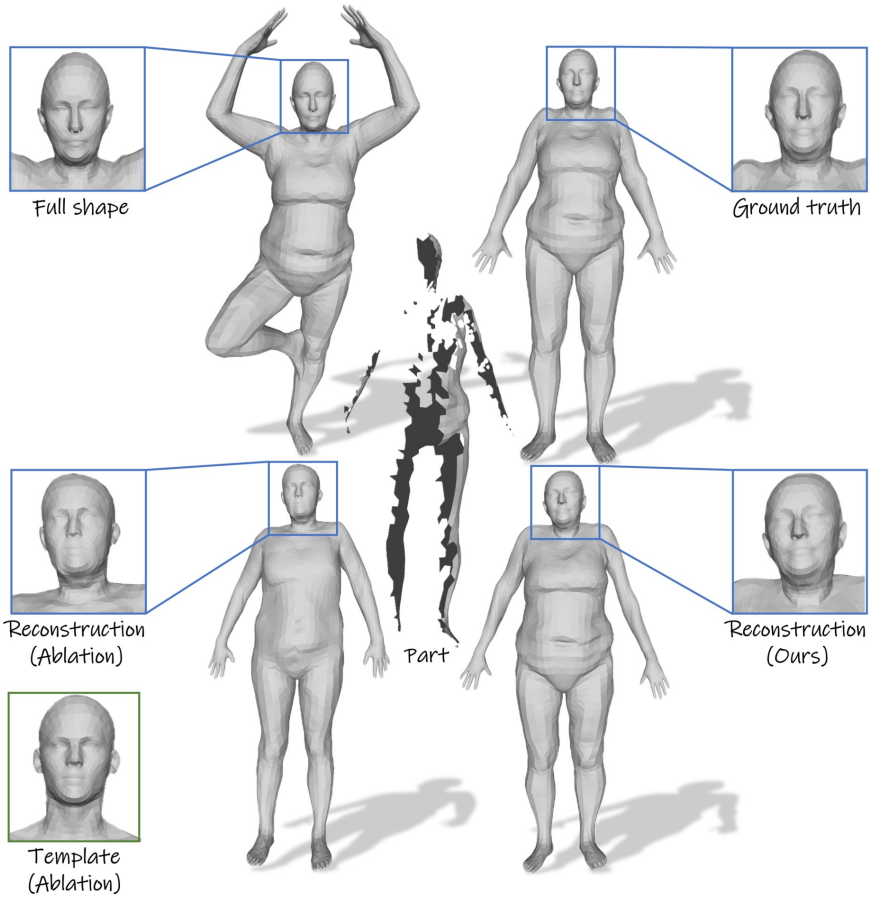

Fig. 2. Comparison with fixed-template ablation experiment. Our method recovers the shape details much more faithfully.

1.2 Robustness Analysis

We now turn to analyze the robustness and stability of our proposed method with regard to three important factors: Noise, sampling, viewpoint and proximity between the source and target poses. For this, we utilize a network trained on the FAUST train set and evaluate over a disjoint test-set of 200 single-view projected scans produced from 10 viewpoints of 2 subjects exhibiting 10 different poses. Each scan P is matched with all possible poses Q of the same subject, achieving a total of 2000 inputs. We utilize a descriptive partial set of the evaluation metrics proposed in section 4.2 of the paper to evaluate each experiment. Lastly, we examine how our network fares when we query for more complex deformations, depicting the averaged reconstruction errors over the azimuthal projection angles as a function of the L2 distance from the template and ground truth shapes.

	Euclidean distance	Volumetric err.	Chamfer GT → Recon.	Chamfer Recon. → GT	Full Chamfer
Ablation	3.74	17.63 ± 7.41	3.00	2.32	5.32
Ours	**2.94**	$\mathbf{7.05 \pm 3.45}$	**2.42**	**1.95**	**4.37**

Table 1. Comparison with Fixed-Template Ablation Experiment. We evaluate our method against an ablation experiment, repeating exactly the same training except of one significant difference: instead of providing the full shape Q_n as described in the main paper, we provided a *constant* full template T in each of the training examples $\{(P_n, T, R_n)\}_{n=1}^{N}$. The template T is used in inference as well, to predict the completion of a given input part P. We report the prediction errors on FAUST test set, while both networks were trained on FAUST train set. The first and second rows summarize the ablation errors and our method errors, respectively.

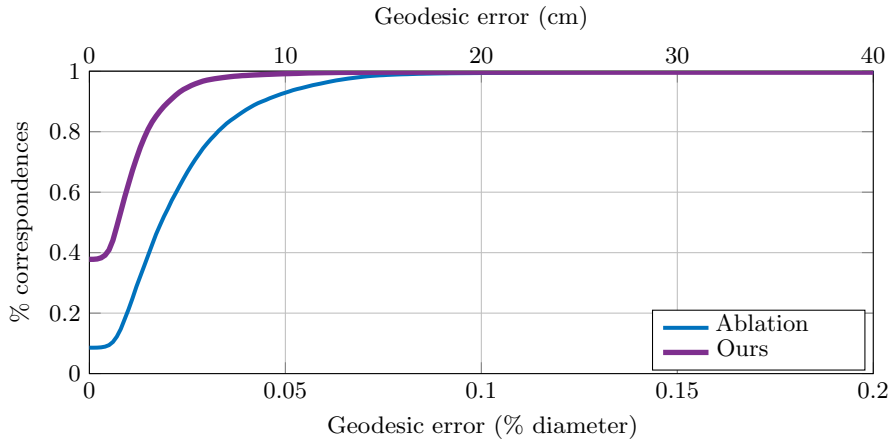

Fig. 3. Comparison with fixed-template ablation experiment. Partial correspondence error evaluated on FAUST Projections dataset.

Residual Noise. In this experiment, we attempt to emulate various artifacts commonly found in real depth scans. We corrupt the vertices of each partial input shape with various degrees of additive white Gaussian noise, with standard deviations in the range [0-4] cm. The corrupted partial shapes are fed to the network, together with the full shapes. Averaged reconstruction statistics are shown in Figure 4 for a Euclidean and the two directional Chamfer distances. As can bee seen, our method accuracy only slightly declines with the increase of the noise.

Downsampling. We evaluate the network performance when provided with a partial shape at a lower resolution. For each partial shape in the test set, we decimate at random some percentage of the existing vertices. As can be seen in Figure 5, even under a majority decimation of the vertices, the proposed network is able to recover well the ground truth shape.

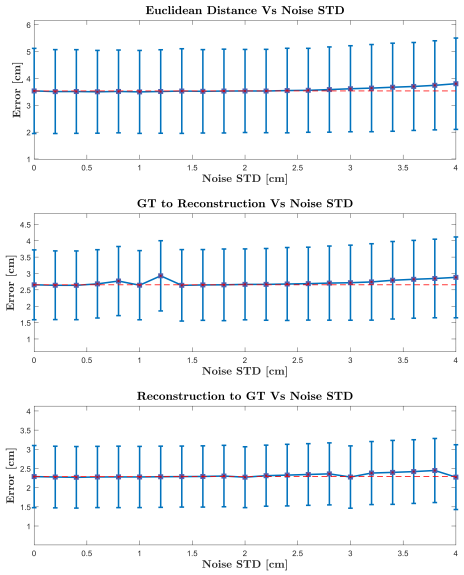

Fig. 4. Robustness to Noise. Three reconstruction metrics evaluated on completions originating from corrupted partial shapes with increasing levels of additive white Gaussian noise. A noise-free reference is marked with a dashed red line.

Projection Angle. Finally, we examine the sensitivity of our network to the projection angle. We note that due to the different projection angles and poses, it is not unreasonable that some angles hold a higher degree of information relevant for reconstruction than others. Ideally, we would like to enable the network a reliable reconstruction at every angle. We partition the 2000 completions received over the test set into their corresponding projection angles, and accumulate the errors over each partition. As seen in Figure 6, the received error distribution is close to uniform, demonstrating an strong resilience to azimuthal change.

Reconstruction quality w.r. template deformation proximity. We examine the reconstruction error as increasingly tougher template shapes are provided. We expect that inserting template shapes that are further from the ground truth will decrease the reconstruction accuracy due to the added problem complexity. We sort the 2000 completions by their vertex wise L2 distance RMSE from the template shape to the ground truth shape, and average the reconstruction errors over all ten azimuthal angles projected from the ground truth shape. As seen in Figure 7, the expected result is received, with the network losing accuracy when we query for more complex deformations.

1.3 Handling Different Types of Partiality

We explored the performance of our network, operating on partial shapes resulting from different types of single view projections. Specifically, we repeated the

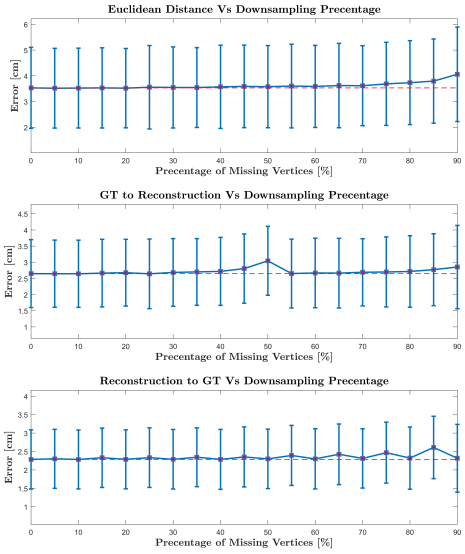

Fig. 5. Robustness to Downsampling. Three reconstruction metrics evaluated on completions originating from decimated partial shapes with varying levels of vertex erasure. A baseline with the evaluation realized with no decimation is marked with a dashed red line.

training on FAUST dataset, sampling the projections uniformly in the azimuthal range of $[0, 360]$ and in the elevation range of $[-90, 90]$. This sampling produced a grid of 10 azimuthal and 11 elevation angles, resulting in 110 different projections per pose. Remarkably, this facilitated an improvement of 10 %, on the solely azimuthally projected test-set reported in the main paper, reducing the error from 2.94 to 2.64 cm. When extending the test-set to infer on all 110 projections, we report an error of 2.78cm, an improvement of 5.4% on a more challenging test-set. We hypothesize this improvement stems from a higher robustness due to the increased data augmentation.

2 Additional Visualizations

Here we provide additional reconstructions that were not included in the main paper in order to save space. Figure 8 and Figure 9 visualize our network predictions for FAUST Projections and AMASS Projections datasets, respectively.

3 Non-Rigid partial correspondence

Figure 10 visualizes the dense correspondence between the input partial and full shape. As explained in the paper, we achieve this by using the network reconstruction as a proxy; For every point in the partial shape we calculate the

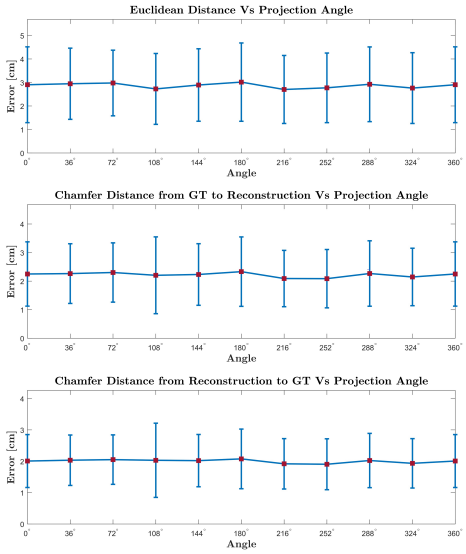

Fig. 6. Robustness to Projection Angle. Three reconstruction metrics evaluated on different groups of the test-set, partitioned by the projection angle. We note a close to uniform distribution over the different angles, attributing to a azimuthal invariancy.

nearest neighbor point in the reconstruction allowing us a recovery of a mapping between the partial shape to the reconstructed shape, which is by construction also the mapping between the part and the full input shape. In Section 4.4 of the paper we evaluated the predicted correspondence numerically for FAUST Projections and AMASS Projections datasets, providing geodesic error graphs for both, in Figure 5 and Figure 6, respectively. For completion, we show the results also qualitatively here.

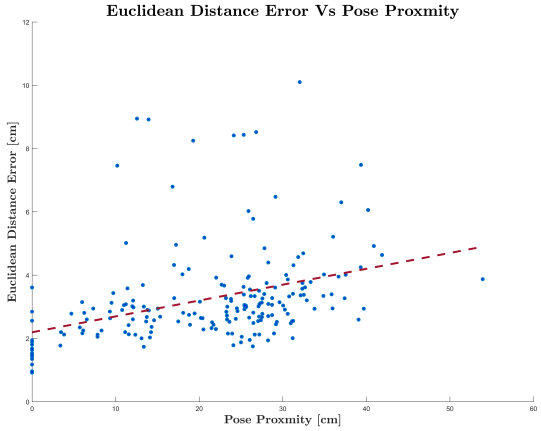

Fig. 7. Euclidean distance as a function of pose proximity. Depiction of the reconstruction error received as a function of the vertex to vertex L2 distance RMSE of the template shape from the ground truth shape. Each dot represents a pair of template and ground truth shapes with the free variable being the distance RMSE between them, and the dependent variable a mean over the reconstruction errors received for the completion with each of the ten azimuthal angles. A linear trend line was fitted to the resulting points for orientation. As can be expected, the greater the deformation distance, the harder it is to deform the template shape to best fit the partial shape.

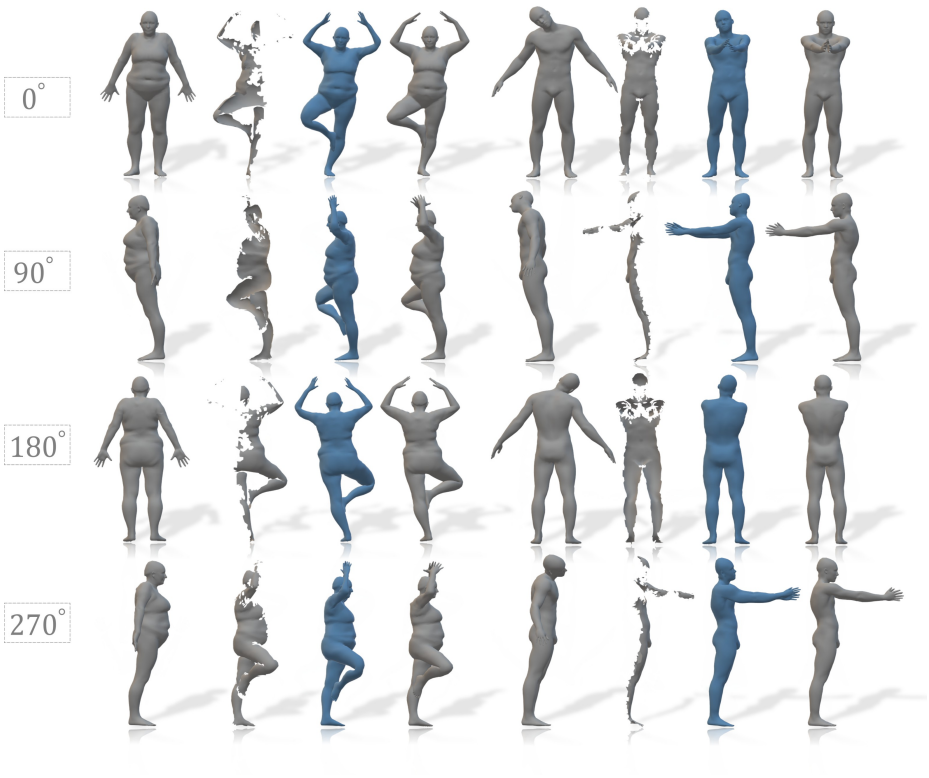

Fig. 8. Predicted completions, FAUST Projections. Each column shows a completion for a different subject, while each row provides a different perspective on the reconstructed 3D model. From left to right: full input shape Q, input part P, predicted completion $F_{\theta(P,Q)}(Q)$, ground truth completion R.

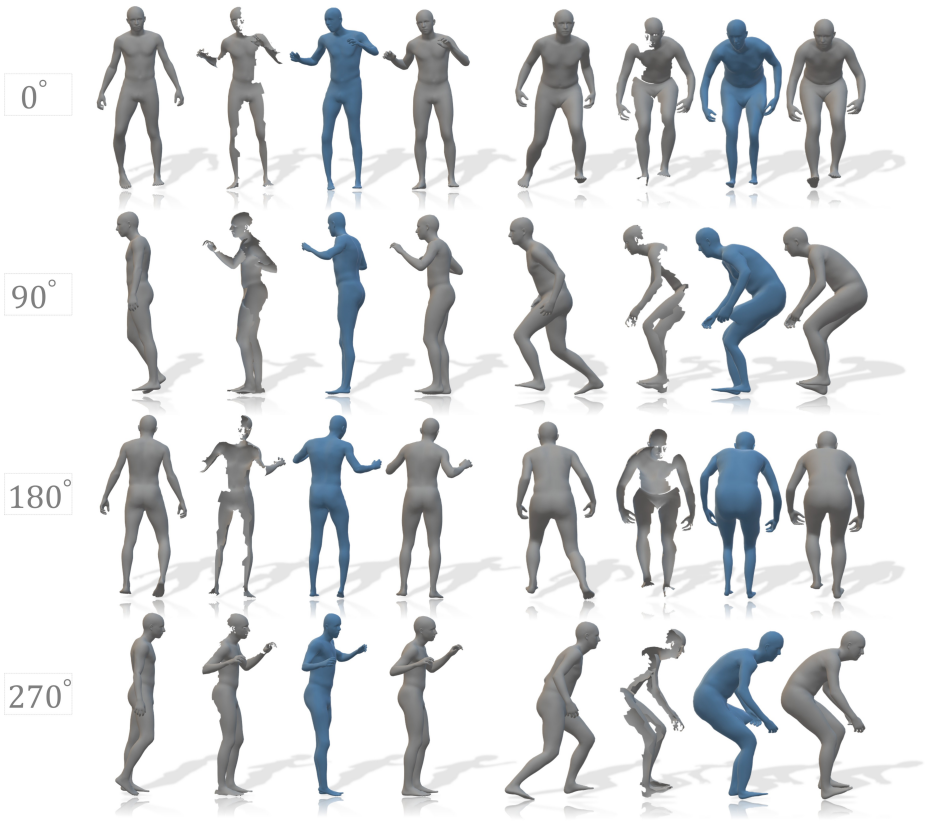

Fig. 9. Predicted completions, AMASS Projections. Each column shows a completion for a different subject, while each row provides a different perspective on the reconstructed 3D model. From left to right: full input shape Q, input part P, predicted completion $F_{\theta(P,Q)}(Q)$, ground truth completion R.

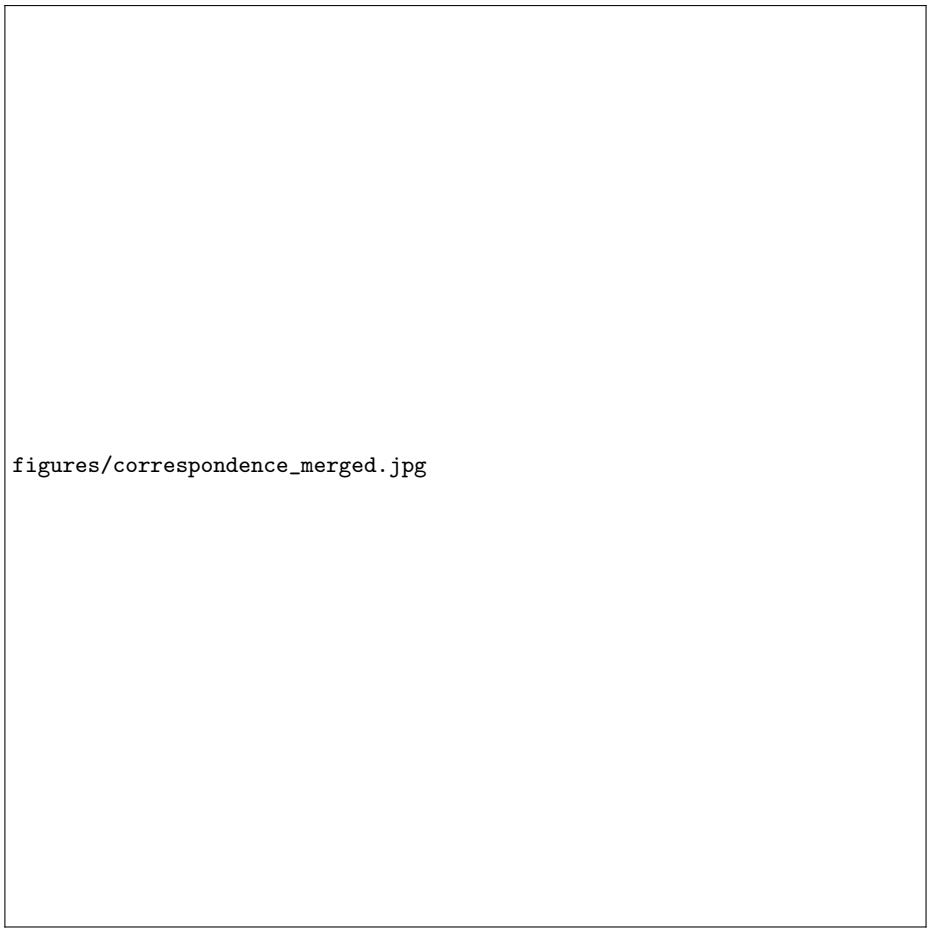

Fig. 10. Non-Rigid partial correspondence. Left and right columns show the dense correspondence for FAUST Projections and AMASS Projections, respectively. From left to right: full input shape Q, our network completion $F_{\theta(P,Q)}(Q)$ and partial input shape P. Corresponding points are indicated by the same color.