Dense Human Body Correspondences Using Convolutional Networks

Lingyu Wei
University of Southern California
cosimo.dw@gmail.com

Qixing Huang
Toyota Technological Institute at Chicago
huangqx@ttic.edu

Duygu Ceylan
Adobe Research
ceylan@adobe.com

Etienne Vouga
University of Texas at Austin
evouga@cs.utexas.edu

Hao Li
University of Southern California
hao@hao-li.com

Appendix I. Comparison

We show that our deep network structure for computing dense correspondences achieves state-of-the-art performance on establishing correspondences between the intra- and inter-subject pairs from the FAUST dataset [1]. For each 3D scan in this dataset, we compute a per-vertex feature descriptor by first rendering depth maps from multiple viewpoints and averaging the per-pixel feature descriptors. Correspondences are then established by nearest neighbor search in the feature space. The accuracy of this direct method is already significantly better than all existing global shape matching methods (that do not require initial poses as input), and is comparable to the state-of-the-art non-rigid registration method proposed by Chen et al. [3], which uses the initial poses of the models to refine correspondences. To make a fair comparison with Chen et al. [3], we use an out-of-the-shelf non-rigid registration algorithm [5] to refine our results. We initialize the registration algorithm with the correspondences established with the nearest-neighbor search and refine their positions after non-rigid alignment. Results obtained with and without this refinement step are reported in Figure 1 and Table 1. It is worth mentioning that per-vertex feature descriptors for each scan are pre-computed. Thus for each pair of scans, we can obtain dense correspondences in less than a second. Though our method is designed for clothed human subjects, our algorithm is far more efficient than all other known methods which rely on local or global geometric properties.

References

[1] F. Bogo, J. Romero, M. Loper, and M. J. Black. FAUST: Dataset and evaluation for 3D mesh registration. In *IEEE CVPR*, 2014.

[2] A. M. Bronstein, M. M. Bronstein, and R. Kimmel. Generalized multidimensional scaling: A framework for isometry-invariant partial surface matching. *Proc. of the National Academy of Science*, pages 1168–1172, 2006.

[3] Q. Chen and V. Koltun. Robust nonrigid registration by convex optimization. In *IEEE ICCV*, 2015.

[4] V. G. Kim, Y. Lipman, and T. Funkhouser. Blended Intrinsic Maps. In *ACM TOG (Siggraph)*, volume 30, 2011.

[5] H. Li, R. W. Sumner, and M. Pauly. Global correspondence optimization for non-rigid registration of depth scans. 2008.

[6] Y. Lipman and T. Funkhouser. Möbius voting for surface correspondence. In *ACM TOG (Siggraph)*, pages 72:1–72:12, 2009.

[7] J. Pokrass, A. M. Bronstein, M. M. Bronstein, P. Sprechmann, and G. Sapiro. Sparse modeling of intrinsic correspondences. In *Computer Graphics Forum*, volume 32, pages 459–468. Wiley Online Library, 2013.

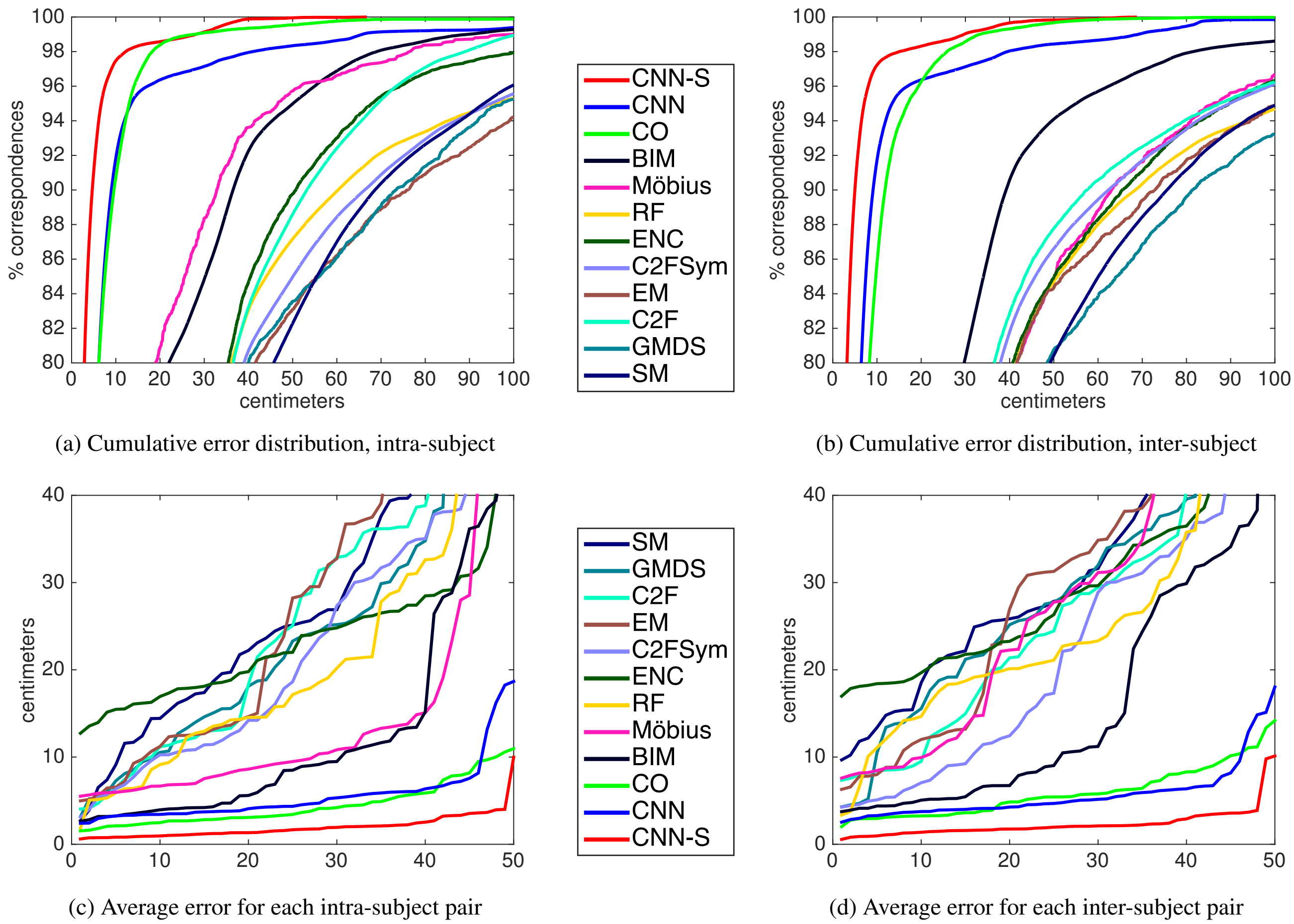
[8] E. Rodola, S. Rota Bulo, T. Windheuser, M. Vestner, and D. Cremers. Dense non-rigid shape correspondence using random forests. 2014.

[9] E. Rodola, A. Torsello, T. Harada, Y. Kuniyoshi, and D. Cremers. Elastic net constraints for shape matching. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 1169–1176, 2013.

[10] Y. Sahillioğlu and Y. Yemez. Coarse-to-fine combinatorial matching for dense isometric shape correspondence. In *Computer Graphics Forum*, volume 30, pages 1461–1470. Wiley Online Library, 2011.

[11] Y. Sahillioğlu and Y. Yemez. Minimum-distortion isometric shape correspondence using em algorithm. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 34(11):2203–2215, 2012.

[12] Y. Sahillioğlu and Y. Yemez. Coarse-to-fine isometric shape correspondence by tracking symmetric flips. In *Computer Graphics Forum*, volume 32, pages 177–189. Wiley Online Library, 2013.



(a) Cumulative error distribution, intra-subject

(b) Cumulative error distribution, inter-subject

(c) Average error for each intra-subject pair

(d) Average error for each inter-subject pair

Figure 1: Evaluation on the FAUST dataset. CNN is the result obtained by performing nearest neighbor search on descriptors produced by our network. CNN-S is the result after non-rigid registration. Data for algorithms other than ours are provided by Chen et al. [3]. Left: Results for intra-subject pairs. Right: Results for inter-subject pairs. Top: Cumulative error distribution for each method, in centimeters. Bottom: Average error for each pair, sorted within each method independently.

method	AE (cm)	worst AE	10cm-recall
CNN-S	**2.00**	**9.98**	**0.975**
CNN	5.65	18.67	*0.918*
CO[3]	4.49	10.96	0.907
RF[8]	13.60	83.90	0.658
BIM[4]	14.99	80.40	0.615
Möbius[6]	22.26	69.26	0.548
ENC[9]	23.60	51.32	0.385
C2FSym[12]	26.87	100.23	0.335
EM[11]	30.11	95.42	0.293
C2F[10]	23.63	73.89	0.334
GMDS[2]	28.94	91.84	0.300
SM[7]	28.81	68.42	0.326

(a) Accuracy on intra-subject pairs

method	AE (cm)	worst AE	10cm-recall
CNN-S	**2.35**	**10.12**	**0.972**
CNN	*5.73*	18.03	*0.917*
CO[3]	5.95	14.18	0.858
RF[8]	17.36	86.76	0.539
BIM[4]	30.58	70.02	0.300
Möbius[6]	26.92	79.43	0.435
ENC[9]	29.29	57.28	0.303
C2FSym[12]	25.89	96.46	0.359
EM[11]	31.25	90.74	0.235
C2F[10]	25.51	90.62	0.277
GMDS[2]	35.06	91.21	0.188
SM[7]	32.66	75.38	0.240

(b) Accuracy on inter-subject pairs

Table 1: Evaluation on the FAUST dataset. CNN is the result obtained by performing nearest neighbor search on descriptors produced by our network. CNN-S is the result after non-rigid registration. Data for algorithms other than ours are provided by Chen et al. [3]. Left: Results for intra-subject pairs. Right: Results for inter-subject pairs. For each method we report the average error on all pairs (AE, in centimeters), the worst average error among all pairs (worst AE), and the fraction of correspondences that are within 10 centimeters of the ground truth (10cm-recall).