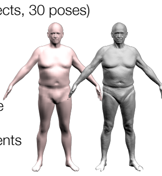
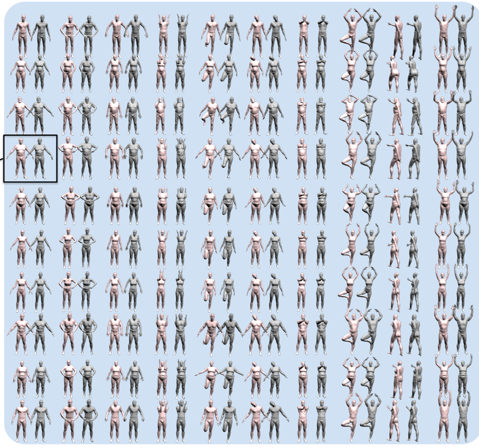
FAUST: Dataset and evaluation for 3D mesh registration

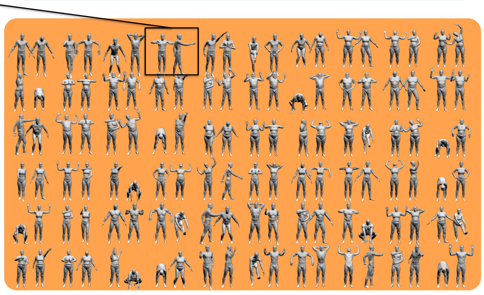
Federica Bogo[1,2], Javier Romero[1], Matthew Loper[1], Michael J. Black[1]
[1]Max Planck Institute for Intelligent Systems, [2]Università degli Studi di Padova

The dataset

- **300** real human body scans (10 subjects, 30 poses)
- **Ground-truth correspondences**: each scan brought into alignment with a common template using a **texture-based registration** technique
- **Training set**: **100** scans + **100** alignments
- **Test set**: **200** scans



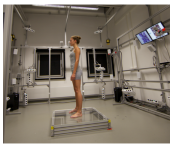




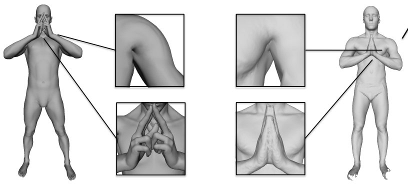
The benchmark

- Intra-subject challenge:
  - 60 scan pairs
  - dense scan-to-scan-correspondences
- Inter-subject challenge:
  - 40 scan pairs
  - sparse scan-to-scan correspondences
- Error metric: average and maximum Euclidean distance between ground truth and provided correspondences

Real vs. synthetic data



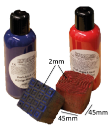
- Each scan is acquired with a high-accuracy 3D multi-stereo system, with 22 RGB cameras for texture capture
- With respect to synthetic datasets (like TOSCA [2]), FAUST scans are much more challenging:
  - realistic deformations
  - missing data
  - different topologies
  - self contacts



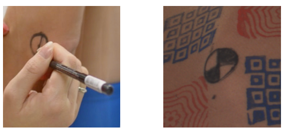
Painted bodies

- Establishing ground-truth correspondences between real scans is difficult
- To achieve accurate registration, we painted the subjects with high-frequency textures

- Intra-subject dense correspondences: high-frequency texture pattern applied with stamps on the subjects' skin

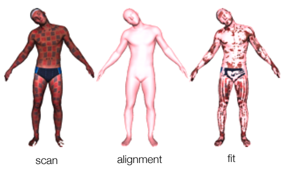


- Inter-subject sparse correspondences: 17 textured markers on specific body points where bones are palpable



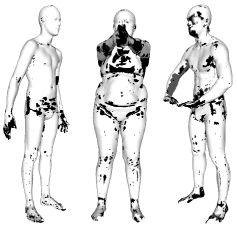
Ground-truth evaluation

- To ensure ground-truth correspondences, we evaluated our alignments in terms of geometry and color:



1. Scan-to-alignment distance: Euclidean distance in 3D space
2. Sliding: optical flow [3] between real and rendered (based on alignments) images

- Scan vertices with too high error for one the metrics are deemed as misaligned (shown in black)
- Main causes of misaligned vertices:
  - missing data (hands, feet)
  - skin stretching
  - clothing



Texture-based registration

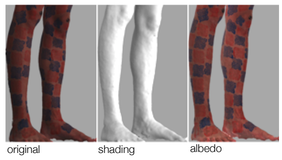
- Given a corpus of scans $\{S^k\}$, we obtain a set of alignments $\{T^k\}$ and learn a set of pose-dependent parameters θ by minimizing:

$$E(\{T^k\},\theta;\{S^k\}) = \lambda_S \sum_k E_S(T^k;S^k) + \lambda_C \sum_k E_C(T^k,\theta;S^k) + \lambda_U \sum_k E_U(T^k;S^k)$$

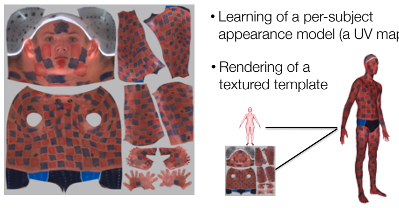
- E_S penalizes distances between mesh surfaces in 3D space
- E_C penalizes deviations from the learned model
- E_U penalizes dissimilarity in **appearance** between scan and template

Appearance-based error term

- Image preprocessing: light estimation and albedo extraction

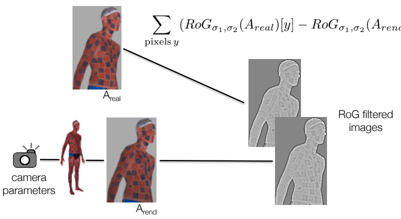


- Learning of a per-subject appearance model (a UV map)
- Rendering of a textured template



- Comparison between real albedo images and rendered images through a robust matching term

$$\sum_{\text{pixels } y} (RoG_{\sigma_1,\sigma_2}(A_{real})[y] - RoG_{\sigma_1,\sigma_2}(A_{rend})[y])^2$$

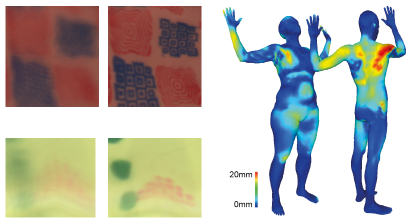


- Error calculation over 22 cameras simultaneously

$$E_U(T^k;S^k) = \sum_{\text{cameras } j} \sum_{\text{pixels } y} (RoG_{\sigma_1,\sigma_2}(A^j_{real})[y] - RoG_{\sigma_1,\sigma_2}(A^j_{rend})[y])^2$$

Benefits

- Texture integrates the incomplete information given by 3D shape in smooth areas (e.g. stomach, torso)



- This results in more accurate intra-subject correspondences, and therefore sharper appearance models

References

[1] F. Bogo, J. Romero, M. Loper, M.J. Black, FAUST: Dataset and evaluation for 3D mesh registration. *CVPR* 2014.
[2] A. Bronstein, M. Bronstein, R. Kimmel, Numerical geometry of non-rigid shapes. *Springer*, 2008.
[3] D. Sun, S. Roth, M.J. Black, A quantitative analysis of current practices in optical flow estimation and the principles behind them. *IJCV*, 106(2):115-137, 2014.