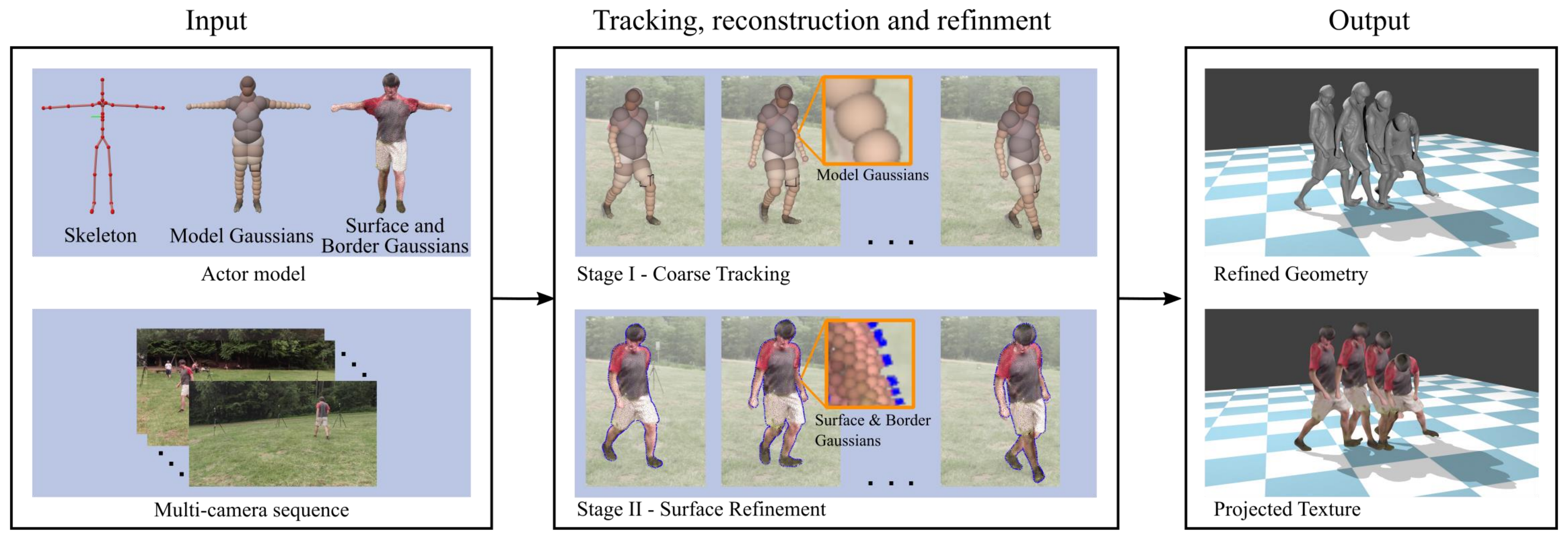
Model-based Outdoor Performance Capture

Nadia Robertini[1,2] Dan Casas[1] Helge Rhodin[1] Hans-Peter Seidel[1] Christian Theobalt[1]
[1]Max Planck Institute for Informatics [2]Intel Visual Computing Institute

Overview



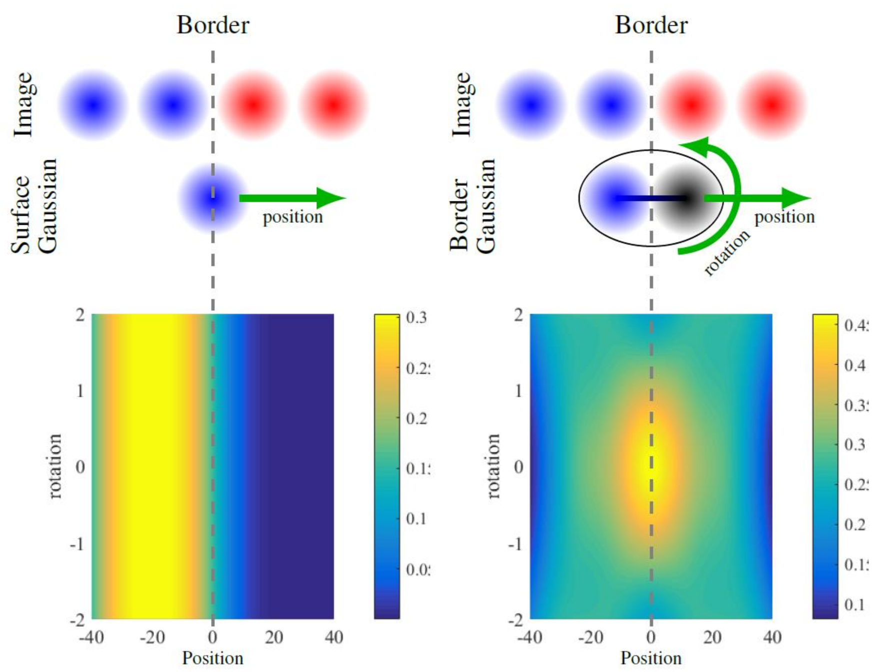
Our goal is to reconstruct human performances captured **outdoors in a multi-camera setup**. We introduce a **unified implicit representation** for both, articulated skeleton tracking and **non-rigid surface shape refinement**. Our method is designed to work on **unsegmented video**.

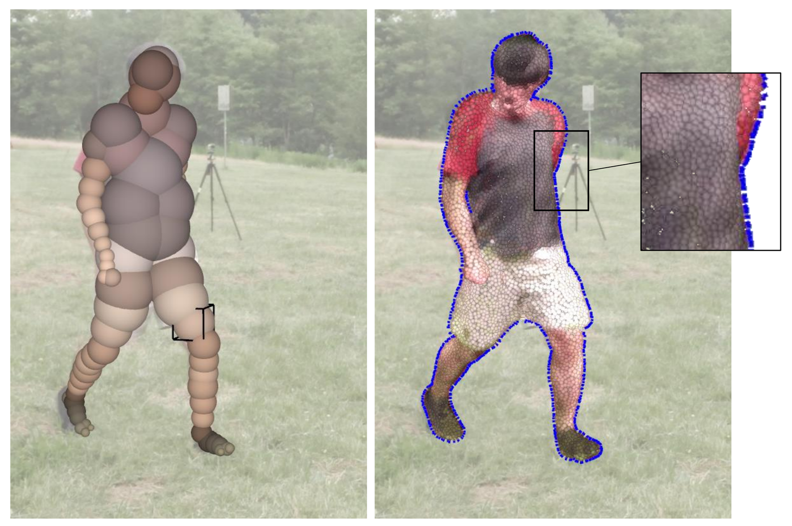
Model Representation

Layer-I approximates the 3D volume of the actor with a collection of colored 3D **Model Gaussians** rigidly attached to the skeleton.

Layer-II is derived from the rigged, colored static surface mesh approximated using two sets of fine 3D Gaussians:

- To optimize the surface interior, **surface Gaussian** are placed at each vertex with the color of the static mesh vertex.
- To align to the surface boundary, **border Gaussians** are placed at the mesh boundary. Inside Gaussians have the surface color and are paired with outside Gaussians with **inverted color similarity**.





Stage-I – Coarse Tracking

The coarse skeleton motion is tracked based on the approach from Stoll et al. [ICCV 2011]

$$E_{m,i} = \left[\int_{\Omega} \hat{g}_m(x)\hat{i}_c(x)\partial x\right]^2 = 2\frac{\sigma_m\sigma_c}{\sigma_m^2+\sigma_c^2}e^{-\frac{\|\mu_m-\mu_c\|^2}{\sigma_m^2+\sigma_c^2}}$$

Stage-II – Surface Refinement

Non-rigid cloth and soft tissue deformation is refined by maximizing the agreement between a fine-scale implicit surface representation and the image (without segmentation).

$$E(\mathbf{v},\boldsymbol{\theta}) = E_{\text{surf}}(\mathbf{v}) + E_{\text{cont}}(\mathbf{v}) - w_{\text{skin}}E_{\text{skin}}(\mathbf{v},\boldsymbol{\theta}) - w_{\text{smooth}}E_{\text{smooth}}(\mathbf{v})$$

Esurf measures the photo consistency of the surface Gaussians with the input images:

$$E_{\text{surf}}(\mathbf{v}) = \sum_{s}^{|\breve{\mathbf{g}}|}\sum_{i}^{|\mathbf{i}|} C(\delta_{s,i})E_{s,i}$$

Econt measures the model-to-image contour alignment, by border Gaussians:

$$E_{\text{cont}}(\mathbf{v}) = \sum_{b}^{|\breve{\mathbf{g}}|}\sum_{i}^{|\mathbf{i}|} C(\delta_{b,i})E_{b_{\text{in}},i} + (1-C(\delta_{b,i}))E_{b_{\text{out}},i}$$

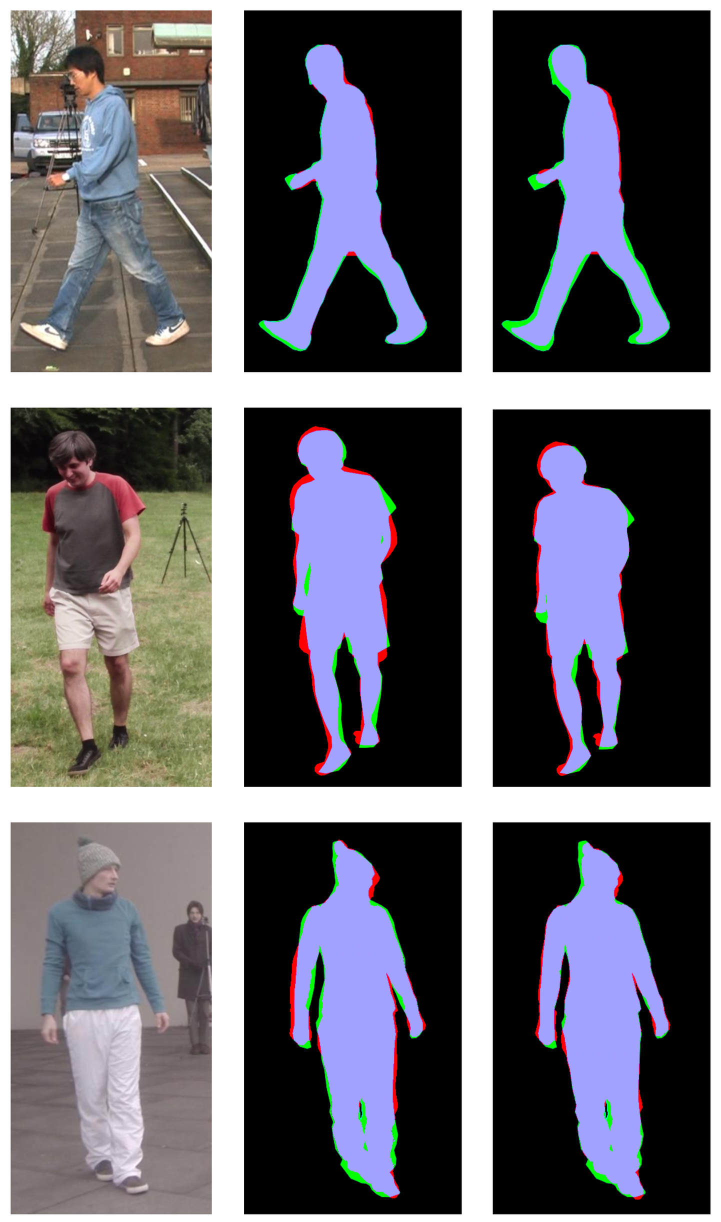
Eskin is a regularization term that maintains the surface attachment to the skeleton and is used to refine the skeletal pose obtained at Stage I.

Esmooth regularizes unnatural surface deformations with a smoothness Laplacian prior term.

Input | Zoom in | Stage-I | Stage-II, no rigidity mask | Stage-II, with rigidity mask

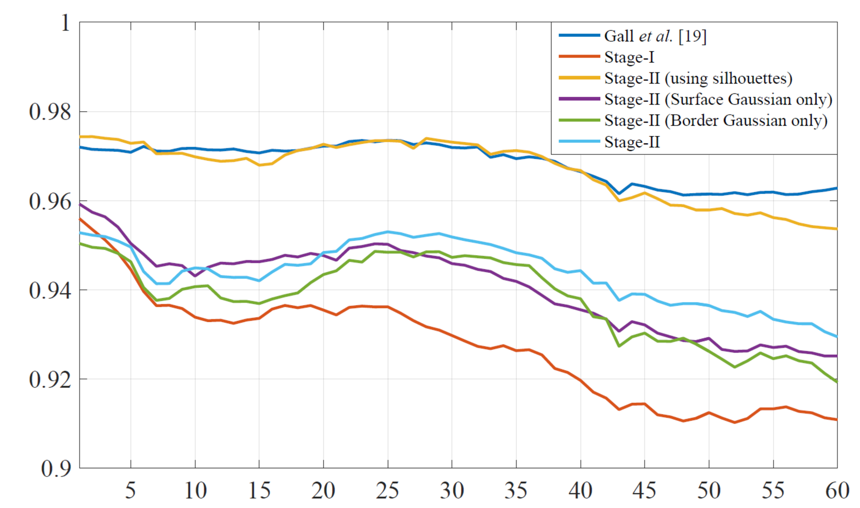
Results and Evaluation

We succeed in outdoor settings with high reconstruction quality, and show that we are on par with state-of-the-art methods on indoor scenes. We quantitatively assess the performance of our method using a silhouette overlap metric.



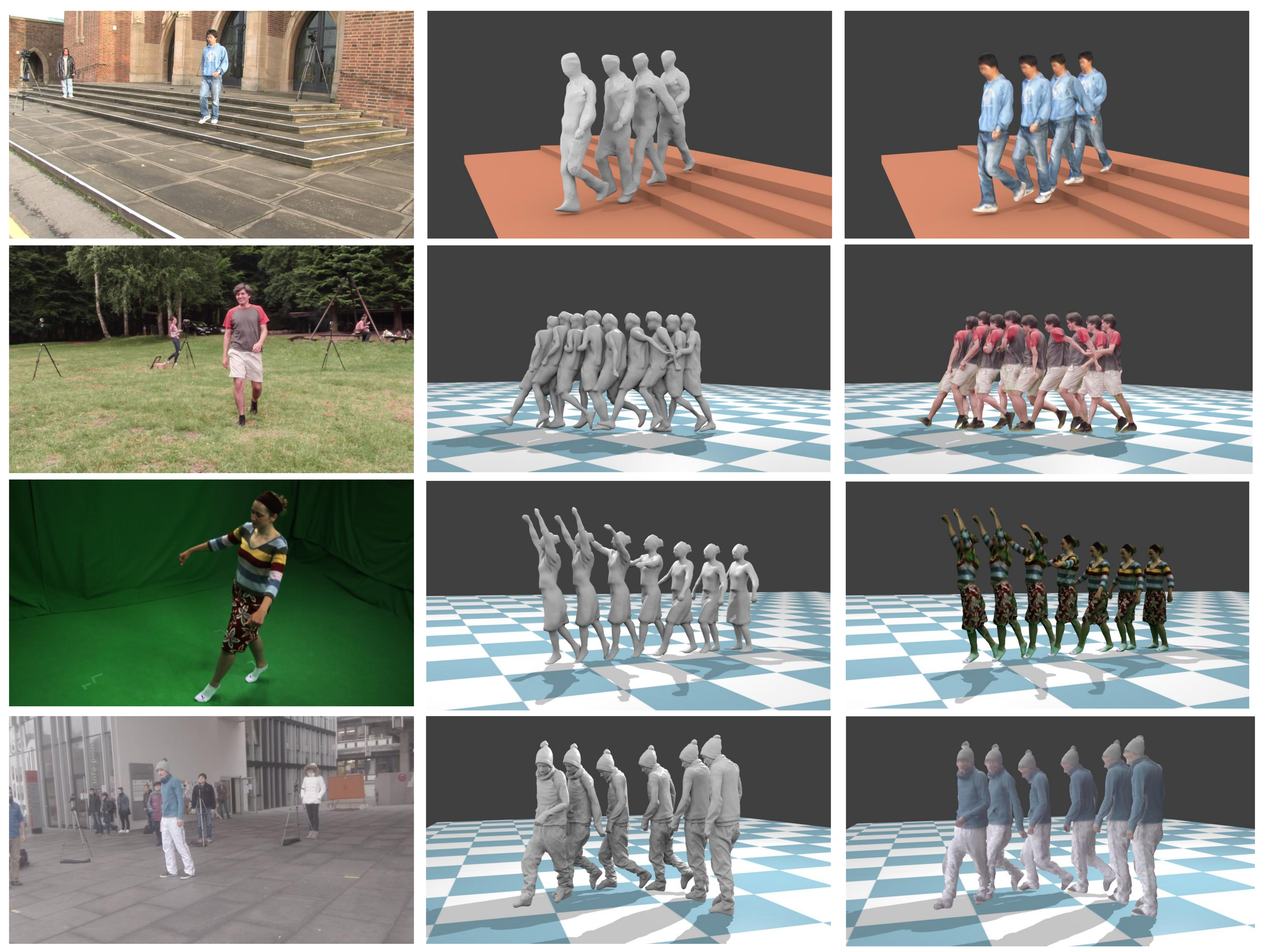
Input | Stage-I | Stage-II

Green: false negative, Red: false positive, Purple: true positive, Black: true negative



	F_1 score	
	Stage-I	Stage-II
cathedral	0.9114 ± 0.0077	0.9362 ± 0.0033
pablo	0.8812 ± 0.0156	0.9212 ± 0.0096
unicampus	0.8962 ± 0.0149	0.9223 ± 0.0083
skirt	0.9271 ± 0.0122	0.9676 ± 0.0056

Table 1: Quantitative evaluation of the sequences tested in this paper. The F_1 score of the Stage-II is consistently higher than in Stage-I.



Funding: ERC starting grant CapReal (335545)
Project page: gvv.mpi-inf.mpg.de/projects/OutdoorPerfcap

N. Robertini, E. De Aguiar, T. Helten, and C. Theobalt. Efficient Multi-view Performance Capture of Fine Scale Surface Detail. 3DV 2014
C. Stoll, N. Hasler, J. Gall, H.-P. Seidel, and C. Theobalt. Fast articulated motion tracking using a sums of Gaussians body model. ICCV 2011
S. Ilic and P. Fua. Implicit meshes for surface reconstruction. PAMI 2006