

General Automatic Human Shape and Motion Capture Using Volumetric Contour Cues

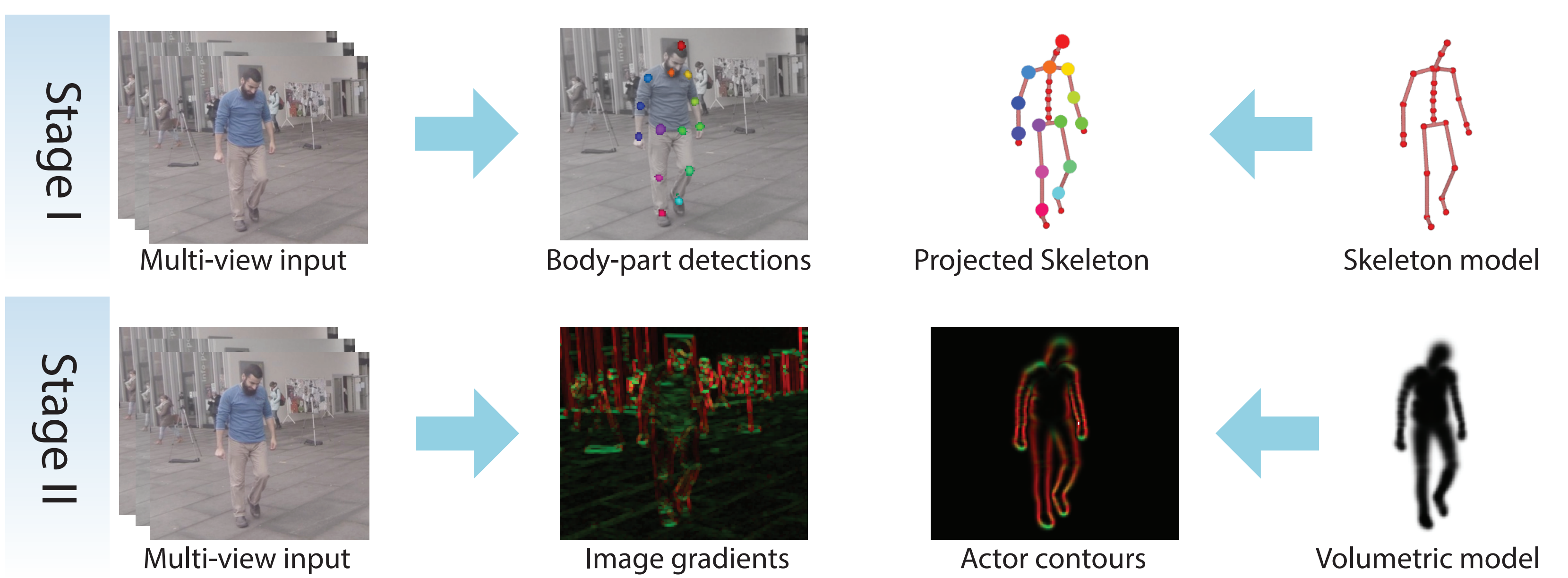
Helge Rhodin¹ Nadia Robertini^{1,2} Dan Casas¹ Christian Richardt^{1,2,3} Hans-Peter Seidel¹ Christian Theobalt¹
¹Max Planck Institute for Informatics ²Intel Visual Computing Institute ³Now at University of Bath

Contributions

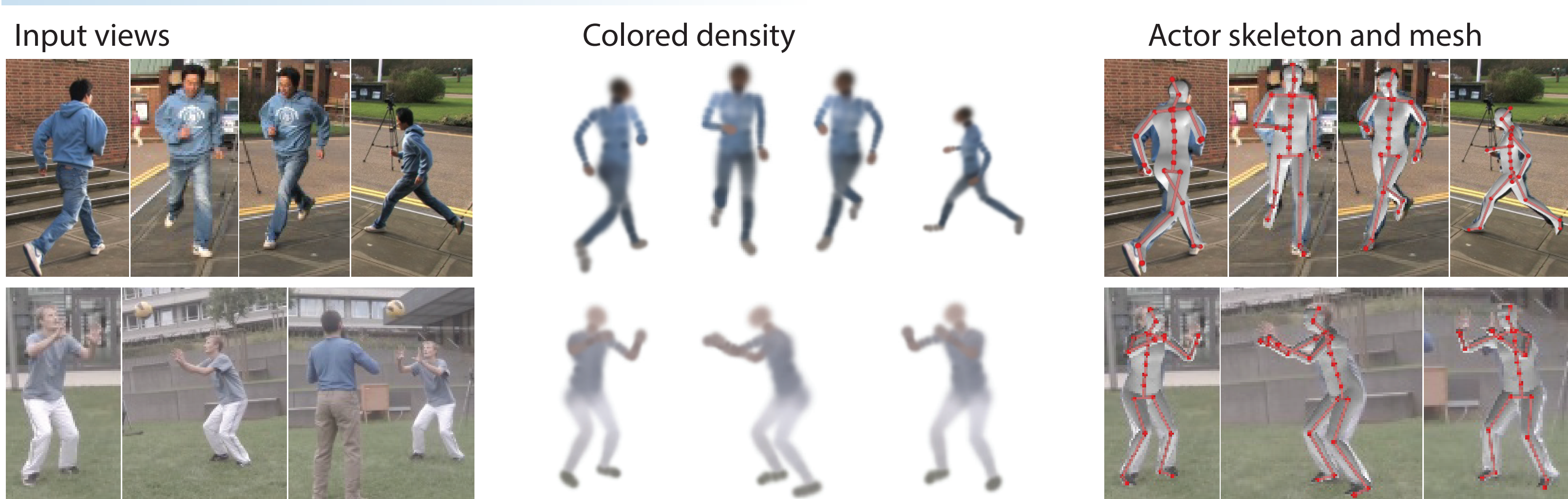
- Fully automatic actor shape and pose estimation in general scenes, unscripted motion
- Joint parametric model volumetric density, mesh, and skeleton
- Spatio-temporal optimization generative & discriminative, top-down & bottom-up

Challenges

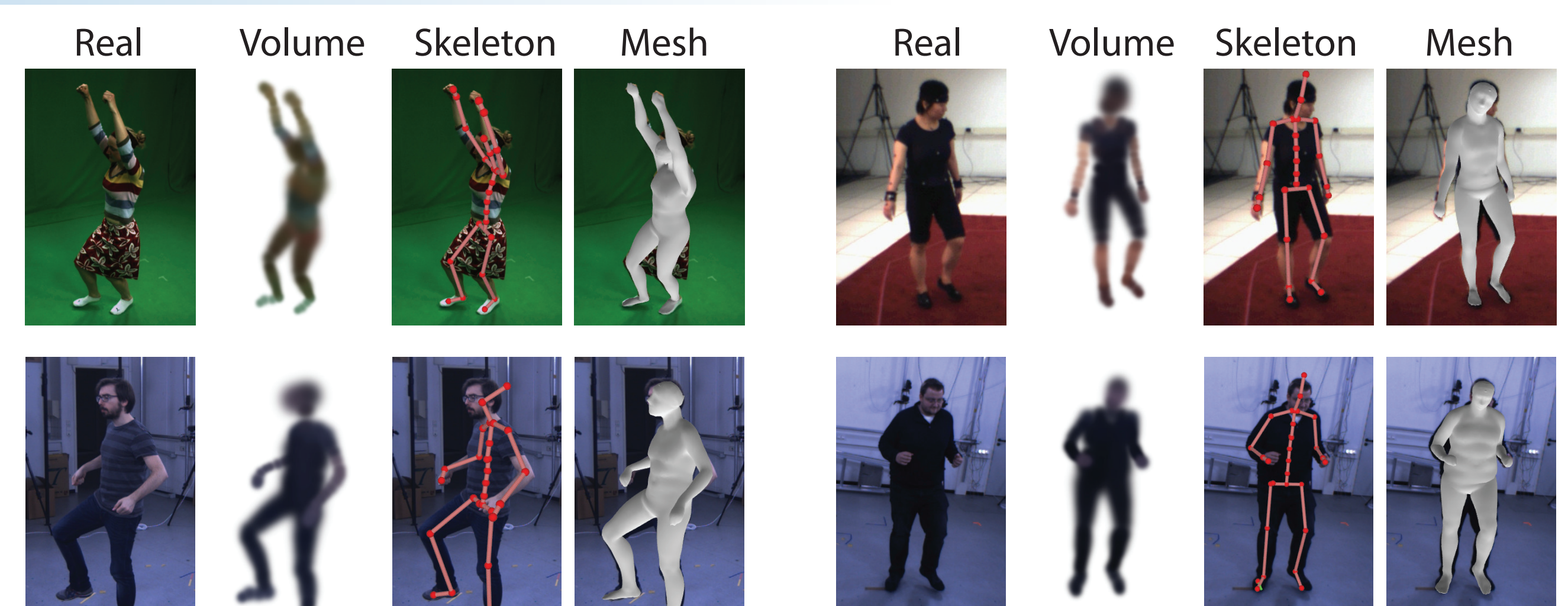
Unknown actor Unknown motion Diverse body shapes Dynamic background Color ambiguities Few views



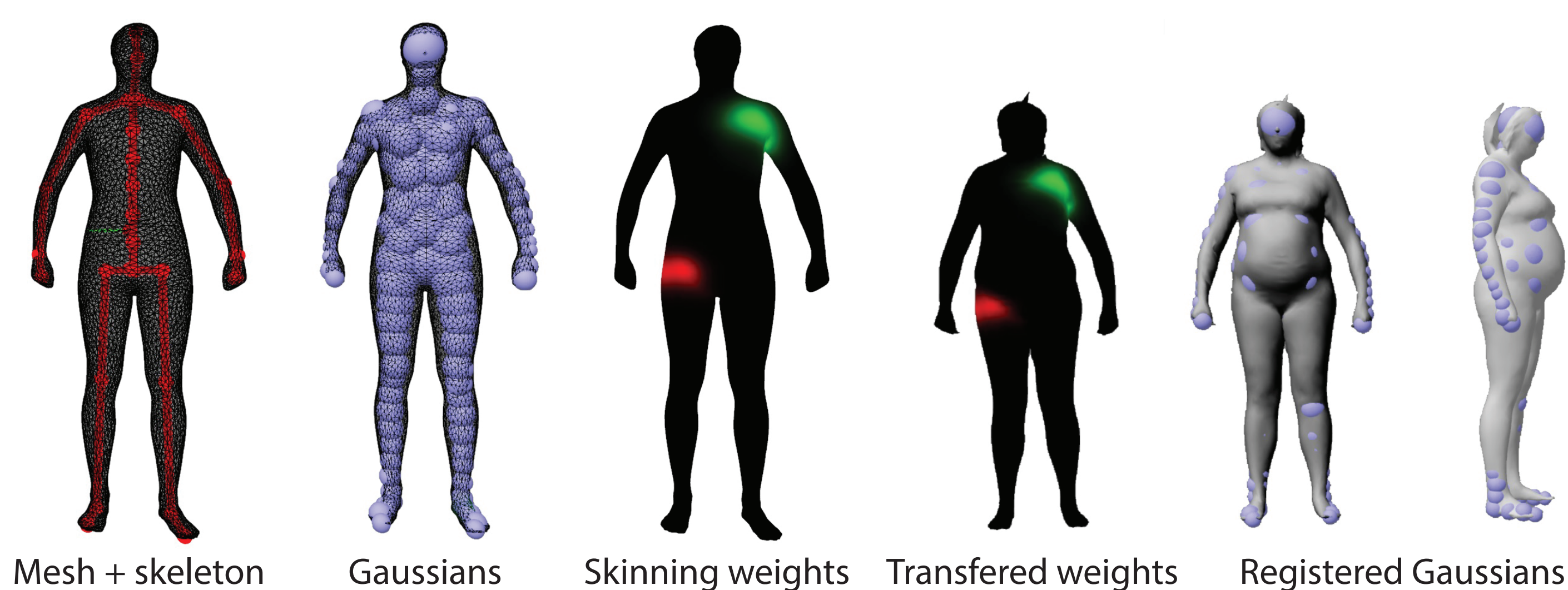
Outdoor reconstructions



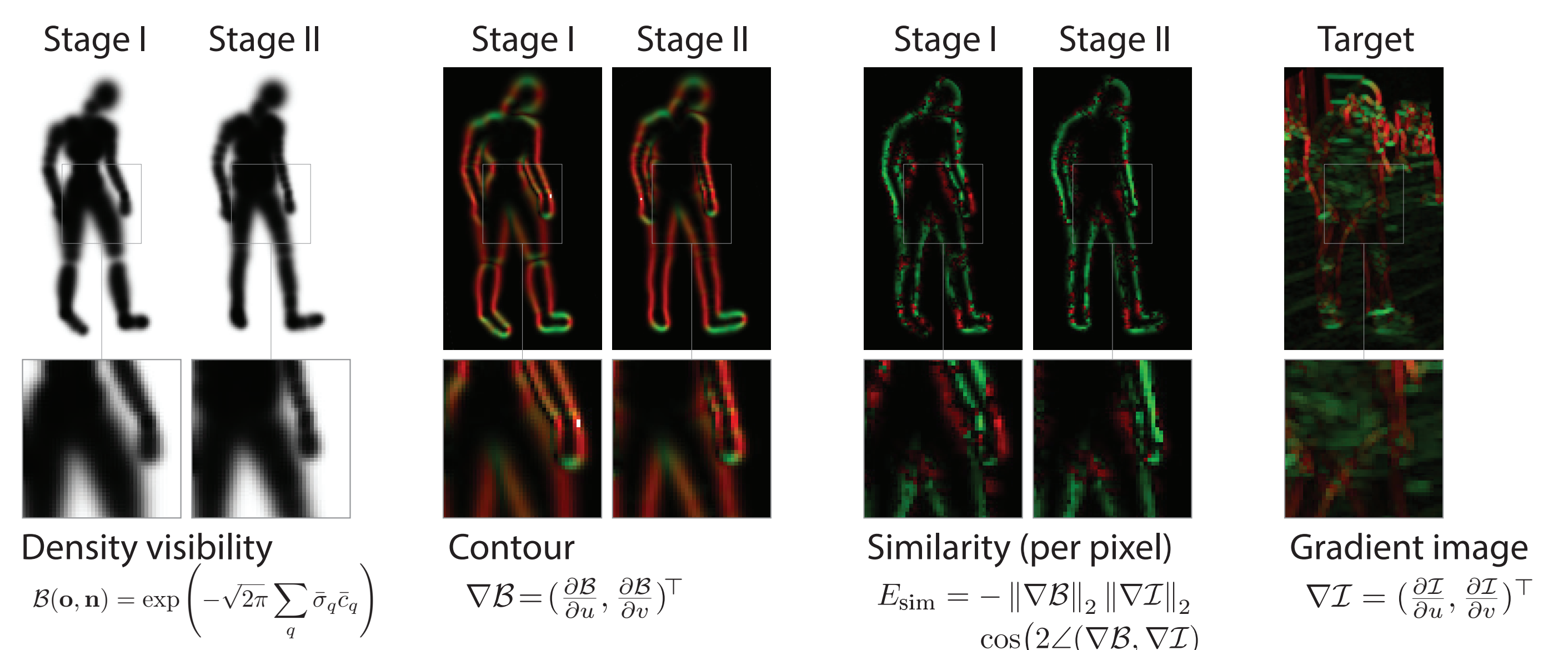
Indoor reconstructions



Parametric actor model



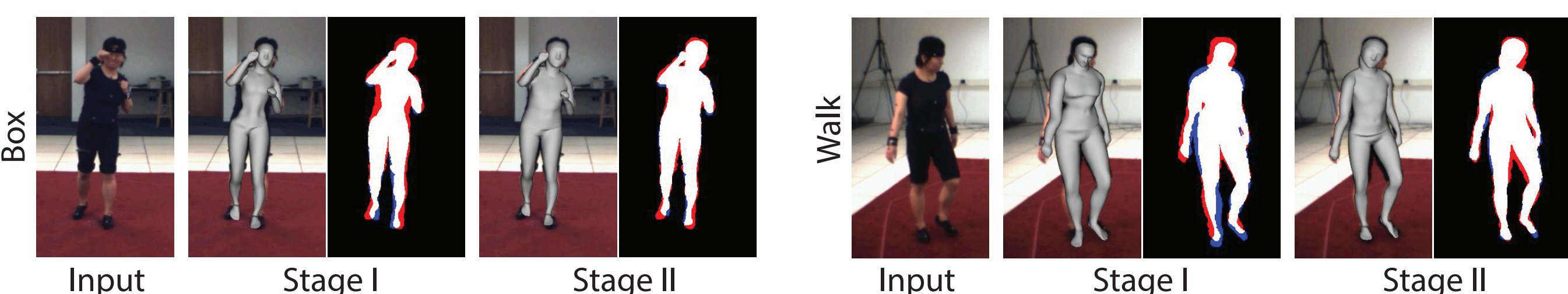
Volumetric contour cues (Stage II)



Evaluation - Shape

	Precision	Recall
Walk Stage I	87.43%	87.25%
Walk Stage II	95.18%	86.89%
Box Stage I	93.26%	81.11%
Box Stage II	95.42%	85.28%

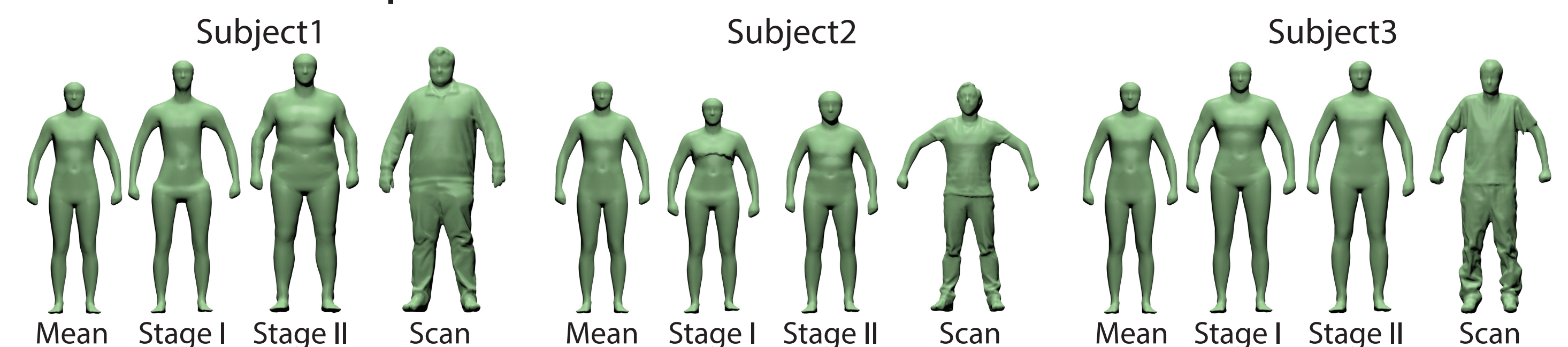
Silhouette overlap metric (Human Eva)



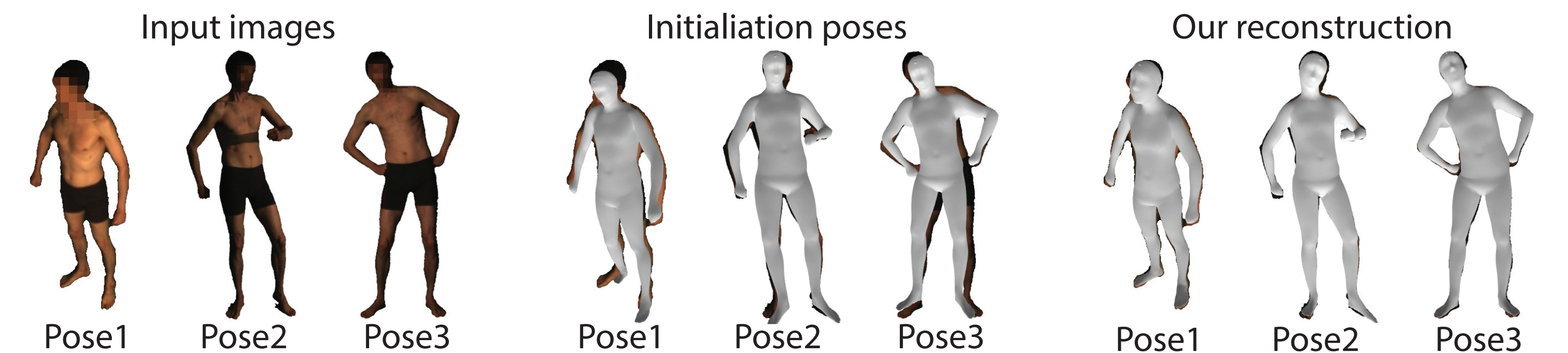
Body dimension metric

	Chest size [cm]				Waist size [cm]				Hip size [cm]				Height [cm]				
	[26]	Stage I	Stage II	GT	[26]	Stage I	Stage II	GT	[26]	Stage I	Stage II	GT	[26]	Stage I	Stage II	GT	
Pose1	92.7	-	92.8	92.6	79.6	-	82.5	80.2	-	-	-	-	183.2	185.0	-	-	-
Pose2	87.4	-	91.3	91.6	78.5	-	82.1	79.4	-	-	-	-	181.9	185.0	-	-	-
Pose3	91.9	-	93.5	91.4	76.9	-	83.2	80.3	-	-	-	-	182.9	185.0	-	-	-
Subject1	-	92.7	132.6	131.3	-	76.7	127.1	132.7	-	108.2	135.4	136.1	-	187.5	194.2	195.0	-
Subject2	-	92.3	99.3	100.1	-	77.3	90.6	96.5	-	92.5	102.3	99.7	-	168.5	162.5	162.0	-

Laser-scan comparison



Monocular estimation

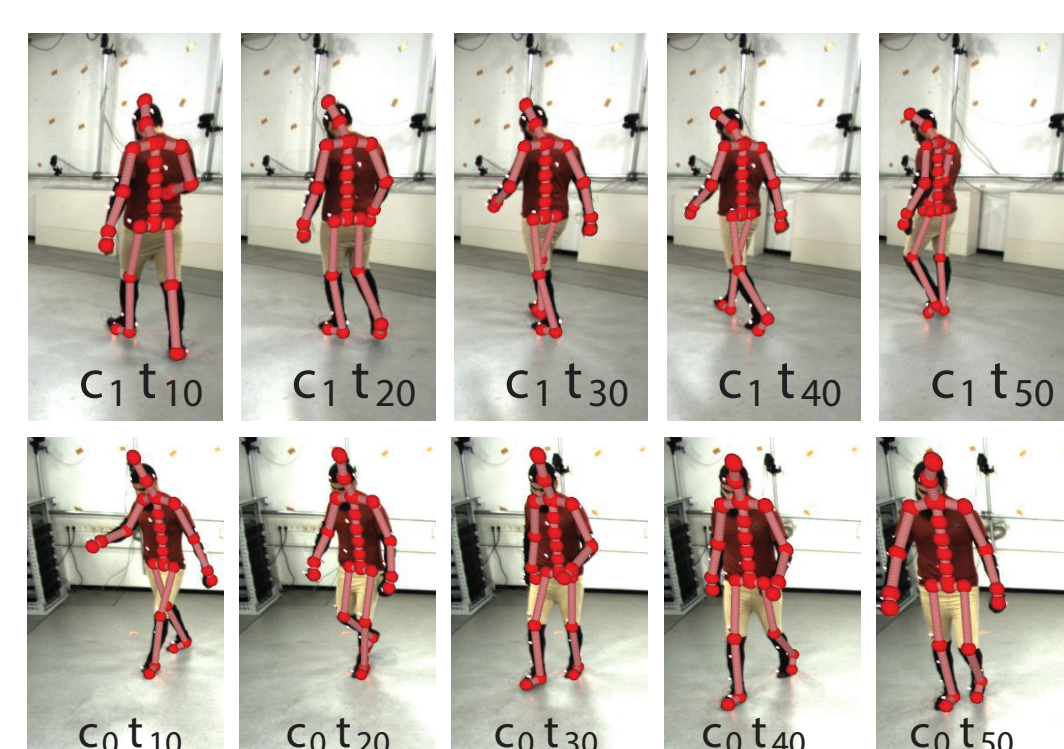


Evaluation - Pose

Human Eva (3D joint position error in mm)

Seq.	Trained on	Our	Amin 2013	Sigal 2012	Belagiannis 2014	Elhayek 2015
S1,	general	74.9 (21.9)	-	66	68.3	66.5
Walk	HumanEva	54.6 (24.2)	54.5	-	-	-
S2,	general	59.7 (15.0)	-	-	62.7	60.0
Box	HumanEva	35.1 (19.0)	47.7	-	-	-

Two-view reconstruction



Related work

Elhayek, A., de Aguiar, E., Jain, A., Tompson, J., Pishchulin, L., Andriluka, M., Bregler, C., Schiele, B., Theobalt, C.: Efficient ConvNet-based marker-less motion capture in general scenes with a low number of cameras. CVPR 2015
 Tompson, J.J., Jain, A., LeCun, Y., Bregler, C.: Joint training of a convolutional network and a graphical model for human pose estimation. NIPS 2014
 Amin, S., Andriluka, M., Rohrbach, M., Schiele, B.: Multi-view pictorial structures for 3D human pose estimation. BMVC 2013
 Belagiannis, V., Amin, S., Andriluka, M., Schiele, B., Navab, N., Ilic, S.: 3D pictorial structures for multiple human pose estimation. CVPR 2014
 Sigal, L., Isard, M., Haussecker, H., Black, M.J.: Loose-limbed people: Estimating 3D human pose and motion using non-parametric belief propagation. IJCV 2012
 Balan, A.O., Sigal, L., Black, M.J., Davis, J.E., Haussecker, H.W.: Detailed human shape and pose from images. CVPR 2007
 Rhodin, H., Robertini, N., Richardt, C., Seidel, H.P., Theobalt, C.: A versatile scene model with differentiable visibility applied to generative pose estimation. ICCV 2015

