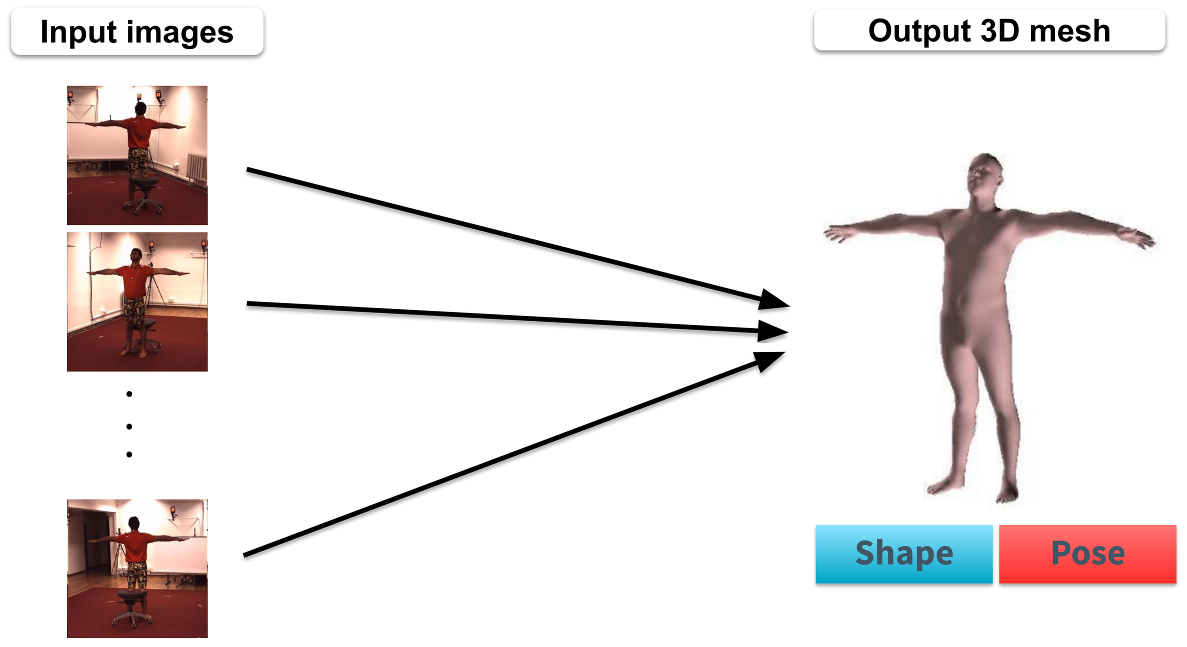
Multi-View Human Model Fitting Using Bone Orientation Constraint and Joints Triangulation

Jordy Ajanohoun Eric Paquette Carlos Vázquez
École de technologie supérieure, Montreal, Quebec, Canada

Introduction

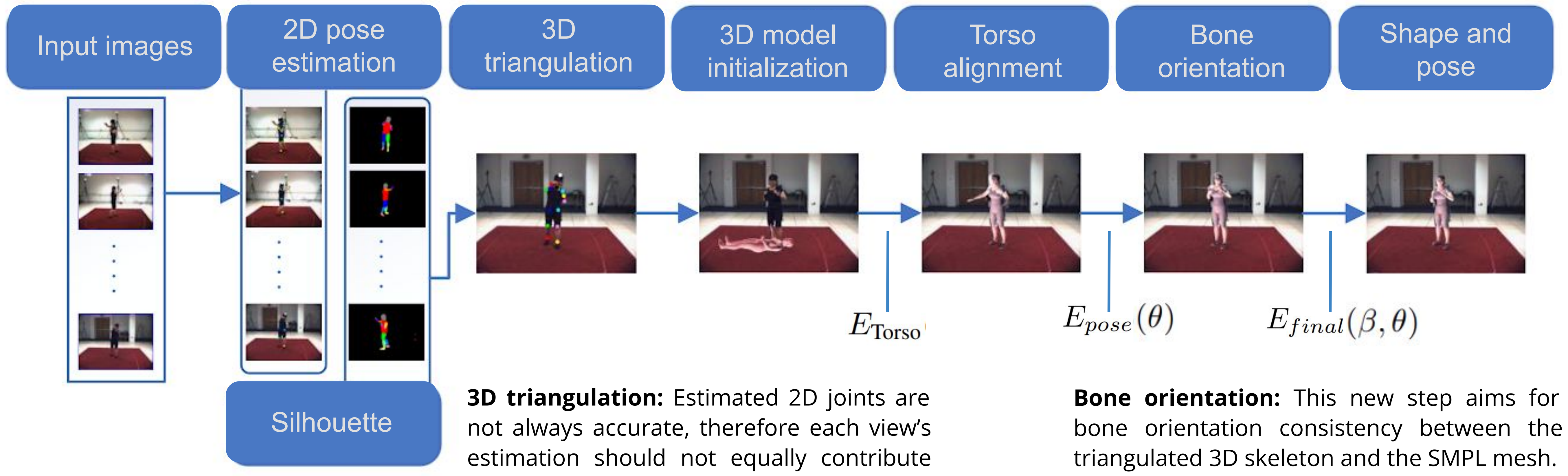
3D human body reconstruction from images consists in generating an accurate 3D mesh of anyone's body only from images.



Methodology

- We make use of the Skinned Multi-Person Linear (SMPL) parametric body model.
- Regress the model parameters that best fit the shape and pose of the individual on the images.

Our approach works as follows:
First, the 2D joints are estimated on each view using a CNN. Then, we use a linear algebraic triangulation to lift estimated 2D joints to 3D, resulting in a joint estimation with fewer errors. Next, the torso of the mesh is aligned with the torso of the individual through the optimization of an energy function. Finally, we fit the mesh to the 3D joints while imposing a bone orientation constraint between the 3D model and the corresponding body parts detected in the images. We do so by minimizing a new set of objective functions through a two-step optimization process that provides a good initialization for the final refinement of the shape (β) and pose (θ) parameters.



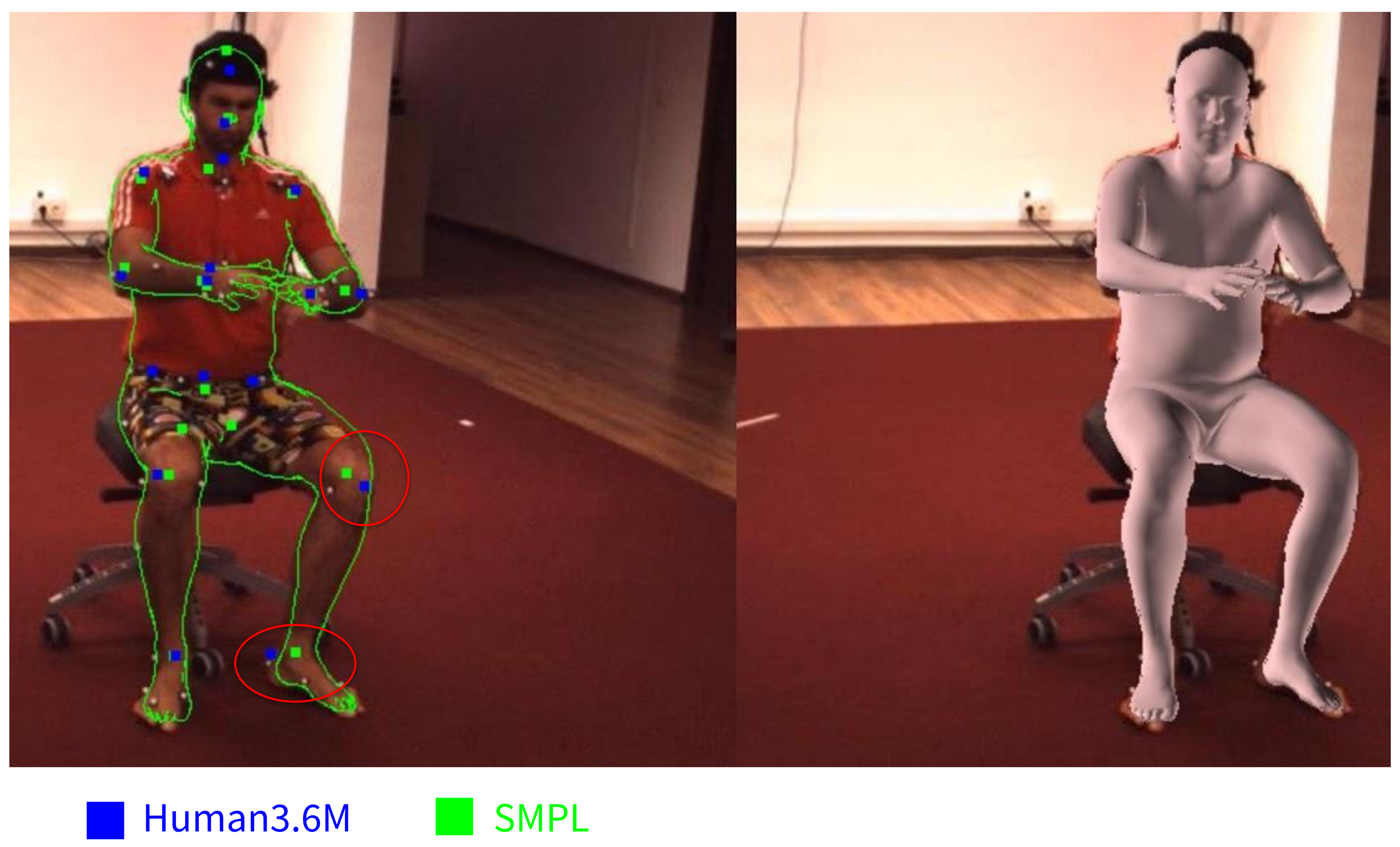
3D triangulation: Estimated 2D joints are not always accurate, therefore each view's estimation should not equally contribute when triangulating. The triangulation is weighted and the weights lead to better estimations for the 3D joints.

Bone orientation: This new step aims for bone orientation consistency between the triangulated 3D skeleton and the SMPL mesh. Our bone orientation constraint (BOC) allows us to decouple pose and shape parameters and to focus on bone orientations.

Evaluation

- The semantic position of joints in the SMPL model and in the Human3.6M data set do not exactly match, despite the fact that the SMPL mesh and its silhouette (green contour) match the individual.
- To account for this discrepancy we introduce, for each joint, a shift vector computed in the joint's local space.
- We measure the performance with the Mean Per Joint Position Error (MPJPE) metric.

Shift between SMPL and Human3.6M joints



Main observations:
- Among the multi-view methods, only MuVS and our approach estimate the shape in addition to the pose. The other methods only focus on 3D joint estimation. This partly explains why the last two methods in the table perform better than our approach.
- Compared to MuVS, our approach is more accurate without using silhouettes or a temporal smoothing stage as MuVS does.
- The BOC and the shift vectors effectively reduce the error in a notable way.

Mean Per Joint Position Error (mm)

Method	Shape	PA	MV	MPJPE
Kanazawa *et al.* (2018)	Yes	Yes	No	66.65
Trumble *et al.* (2018)	No	No	Yes	62.50
Kolotouros *et al.* (2019a)	Yes	Yes	No	62.00
Pavlakos *et al.* (2017b)	No	No	Yes	56.89
MuVS$^{\text{S, T}}$	Yes	Yes	Yes	47.09
Ours	Yes	Yes	Yes	54.86
Ours$^{\text{SV}}$	Yes	Yes	Yes	39.56
Ours$^{\text{BOC}}$	Yes	Yes	Yes	46.37
Ours$^{\text{BOC, SV, S}}$	Yes	Yes	Yes	33.07
Ours$^{\text{BOC, SV}}$	Yes	Yes	Yes	30.13
Iskakov *et al.* (2019)	No	Yes	Yes	20.80
He *et al.* (2020)	No	Yes	Yes	19.00

S: Silhouettes T: Temporal SV: Shift vectors BOC: Bone orientation step

Conclusion

We summarize the main contributions of our approach as follows:
- A bone orientation constraint (BOC) to recover the pose parameter independently from the shape parameter
- A more precise initialization for the simultaneous optimization of pose and shape parameters thanks to the BOC
- A two-step optimization process that improves the accuracy of the pose and shape estimations
- The shift vectors

References

M. Loper, N. Mahmood, J. Romero, G. Pons-Moll, and M. J. Black, “SMPL: a skinned multi-person linear model”, *ACM Trans. Graph.*, vol. 34, no. 6, pp. 1-16, 2015.

K. Iskakov, E. Burkov, V. Lempitsky, and Y. Malkov, “Learnable Triangulation of Human Pose”, in *IEEE ICCV*, 2019, pp. 7717-7726.

Y. Huang, F. Bogo, C. Lassner, A. Kanazawa, P. V. Gehler, J. Romero, I. Akhter, and M. J. Black, “Towards Accurate Marker-Less Human Shape and Pose Estimation over Time”, in *3DV*, 2017, pp. 421-430.

F. Bogo, A. Kanazawa, C. Lassner, P. Gehler, J. Romero, and M. J. Black, “Keep It SMPL: Automatic Estimation of 3D Human Pose and Shape from a Single Image”, in *ECCV*. Springer, 2016, vol. 9909, pp.561-578.

C. Ionescu, D. Papava, V. Olaru, and C. Sminchisescu, “Human3.6m: Large scale datasets and predictive methods for 3d human sensing in natural environments”, *IEEE TPAMI*, 2014.