Better Prior Knowledge Improves Human-Pose-Based Extrinsic Camera Calibration

Olivier Moliner [1,2] Sangxia Huang [2] Kalle Åström [1]

[1]Lund University, Sweden [2]Sony R&D Center Lund, Sweden

Motivation

Problem statement

- **Accurate extrinsic calibration of wide baseline multi-camera systems** with classical Structure-from-Motion methods requires special calibration equipment and trained operators.
- This is costly and time-consuming, and limits the ease of adoption of multi-camera 3D scene analysis technologies.

Prior work

- Use human pose estimation models to establish point correspondences, thus removing the need for any special equipment [5, 6].
- Challenge: human pose estimation algorithms produce much less accurate feature points compared to patch-based methods.

Our contribution

We introduce several novel ideas to improve the accuracy of human-pose-based extrinsic calibration. In particular:

- **A robust reprojection loss** more suitable for camera calibration with human poses.
- We introduce a **3D-human-pose likelihood model** to the objective function of bundle adjustment.

References

[1] Luca Ballan and Guido Maria Cortelazzo. Marker-less motion capture of skinned models in a four camera set-up using optical flow and silhouettes. In *3DPVT*, Atlanta, GA, USA, June 2008.

[2] Cristian Sminchisescu Catalin Ionescu, Fuxin Li. Latent structured models for human pose estimation. In *International Conference on Computer Vision*, 2011.

[3] Catalin Ionescu, Dragos Papava, Vlad Olaru, and Cristian Sminchisescu. Human3.6m: Large scale datasets and predictive methods for 3d human sensing in natural environments. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 36(7):1325--1339, jul 2014.

[4] Hanbyul Joo, Tomas Simon, Xulong Li, Hao Liu, Lei Tan, Lin Gui, Sean Banerjee, Timothy Scott Godisart, Bart Nabbe, Iain Matthews, Takeo Kanade, Shohei Nobuhara, and Yaser Sheikh. Panoptic studio: A massively multiview system for social interaction capture. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2017.

[5] Jens Puwein, Luca Ballan, Remo Ziegler, and Marc Pollefeys. Joint camera pose estimation and 3D human pose estimation in a multi-camera setup. In *Computer Vision -- ACCV 2014*, pages 473--487. Springer International Publishing, November 2014.

[6] Kosuke Takahashi, Dan Mikami, Mariko Isogawa, and Hideaki Kimata. Human pose as calibration pattern; 3D human pose estimation with multiple unsynchronized and uncalibrated cameras. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pages 1775--1782, 2018.

Method

Preprocessing: Extract body joints from synchronized video streams using human pose estimation.

Initial estimation of the camera extrinsics and 3D positions of the body joints following standard SfM approaches with the body joints as point correspondences between cameras.

Bundle adjustment: To refine the initial estimates, we minimize an objective function consisting of a modified reprojection error and prior models on the plausibility of the estimated motion and poses:

$$E = E_{rep} + E_{motion} + E_{limb} + E_{KCS}$$

- **Robust Reprojection Error:**

$$E_{rep} = \frac{1}{\sum_{i,j,t} w_{ijt}} \sum_{i,j,t} w_{ijt} L(u_{ijt}, \pi_{\mathbf{i}}(\mathbf{U_{jt}}))\,,$$

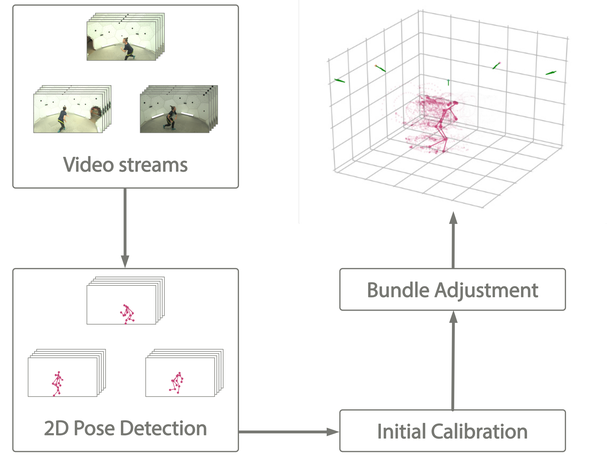
where $L(\cdot,\cdot)$ is the Huber loss function, and the weights w_{ijt} depend on the joint detection scores and the distances between the joints and the cameras.

- **Motion Prior:** E_{motion} is the l_2-norm of the fourth-order derivative of the joint positions, to encourage smooth joint trajectories while accounting for complex human motion.
- **Constant Limb Length Constraint:** E_{limb} enforces the reconstructed limb lengths to stay constant throughout the whole sequence.
- **Body Pose Prior:** To encourage the reconstruction of plausible human poses, E_{KCS} is the average likelihood of the 3D human poses, given by a PCA model fitted on the Human 3.6M dataset.

Conclusion

We introduced several ideas in this paper and achieved improved accuracy for extrinsic camera calibration using human body joints.

We showed that robust loss functions and relevant prior models are effective in handling errors in human body joint detection.



Experimental Results

We evaluate our algorithm on four datasets for which ground truth camera calibration is available: Human 3.6M [2, 3], CMU Panoptic [4], and the Soccer Juggling and Sword Swing sequences [1].

	Puwein et al. [5]		Proposed Solution	
	Pos.	Ang.	Pos.	Ang.
Soccer	5.0	1.0	1.7	0.4
Sword	5.8	1.0	0.9	0.4

Table 1:Comparing our proposed solution compared to [5] on the Soccer and Sword sequences.

Ablation study

ID	Reproj.	Motion	KCS	Limb	H36M `Walking` Pos.	H36M `Walking` Ang.	H36M `WalkTogether` Pos.	H36M `WalkTogether` Ang.	Dance Pos.	Dance Ang.	Soccer Pos.	Soccer Ang.	Sword Pos.	Sword Ang.
0	Initial calibration				4.41 ± 2.66	0.54 ± 0.20	5.81 ± 3.25	0.67 ± 0.34	5.56 ± 1.21	0.78 ± 0.24	13.84 ± 3.86	3.52 ± 1.20	19.86 ± 2.48	4.21 ± 0.45
1					4.87 ± 1.50	0.60 ± 0.15	4.11 ± 1.52	0.53 ± 0.20	3.89 ± 0.36	0.54 ± 0.03	3.47 ± 0.05	0.60 ± 0.01	2.46 ± 0.12	1.20 ± 0.00
2	✓				2.04 ± 0.77	0.31 ± 0.09	2.88 ± 2.08	0.36 ± 0.22	4.17 ± 0.51	0.49 ± 0.16	1.87 ± 0.09	0.47 ± 0.01	1.10 ± 0.12	0.38 ± 0.02
3	✓	✓			2.04 ± 0.77	0.31 ± 0.09	2.84 ± 2.00	0.35 ± 0.21	4.05 ± 0.44	0.46 ± 0.14	2.04 ± 0.14	0.49 ± 0.02	1.09 ± 0.11	0.38 ± 0.02
4	✓		✓		1.88 ± 0.71	0.29 ± 0.09	2.60 ± 1.85	0.33 ± 0.19	4.04 ± 0.44	0.47 ± 0.15	2.10 ± 0.11	0.49 ± 0.02	1.00 ± 0.08	0.37 ± 0.02
5	✓			✓	2.00 ± 0.76	0.31 ± 0.09	2.85 ± 2.32	0.37 ± 0.25	4.09 ± 0.45	0.46 ± 0.14	1.44 ± 0.09	0.43 ± 0.02	0.89 ± 0.09	0.38 ± 0.01
6	✓	✓		✓	1.96 ± 0.74	0.30 ± 0.09	2.81 ± 2.25	0.36 ± 0.24	4.01 ± 0.40	0.45 ± 0.13	1.80 ± 0.12	0.48 ± 0.02	0.89 ± 0.08	0.38 ± 0.01
7		✓	✓	✓	4.36 ± 1.07	0.53 ± 0.11	4.21 ± 1.62	0.52 ± 0.20	4.13 ± 0.53	0.51 ± 0.11	2.16 ± 0.24	0.70 ± 0.02	2.44 ± 0.12	1.00 ± 0.01
8	✓	✓	✓	✓	1.89 ± 0.72	0.29 ± 0.09	2.66 ± 2.08	0.34 ± 0.22	4.02 ± 0.42	0.45 ± 0.14	1.66 ± 0.12	0.44 ± 0.02	0.86 ± 0.05	0.38 ± 0.01
9	Plain vanilla BA with $\theta_{huber} = 0.7$				2.68 ± 0.79	0.33 ± 0.09	2.81 ± 1.17	0.35 ± 0.13	4.16 ± 0.65	0.46 ± 0.09	2.62 ± 0.09	0.69 ± 0.01	1.32 ± 0.02	0.91 ± 0.00
10	Our solution with $\theta_{huber} = 0.7$				2.00 ± 0.76	0.31 ± 0.09	2.69 ± 2.09	0.35 ± 0.22	4.03 ± 0.46	0.46 ± 0.15	1.50 ± 0.10	0.42 ± 0.02	0.96 ± 0.07	0.39 ± 0.02

Table 2:Ablation study