Skeleton Driven Non-rigid Motion Tracking and 3D Reconstruction

Shafeeq Elanattil[1,2], Peyman Moghadam[1,2], Simon Denman[2], Sridha Sridharan[2], Clinton Fookes[2]

[1]Autonomous Systems Laboratory, CSIRO Data61, Brisbane, Australia
[2]Queensland University of Technology, Brisbane, Australia

www.data61.csiro.au

We presents a method for reconstructing the non-rigid surface motion of human performance using a moving RGB-D camera. Current approaches use local optimization and needs many iterations to converge and may get stuck in local minima during sudden articulated movements. We propose a puppet model-based tracking approach using skeleton prior, which provides a better initialization for tracking articulated movements. The proposed approach uses an aligned puppet model to estimate correct correspondences for human performance capture. We also contribute a synthetic dataset which provides ground truth locations for frame-by-frame geometry and skeleton joints of human subjects. Experimental results show that our approach is more robust when faced with sudden articulated motions, and provides better 3D reconstruction compared to the existing state-of-the-art approaches.

Methodology

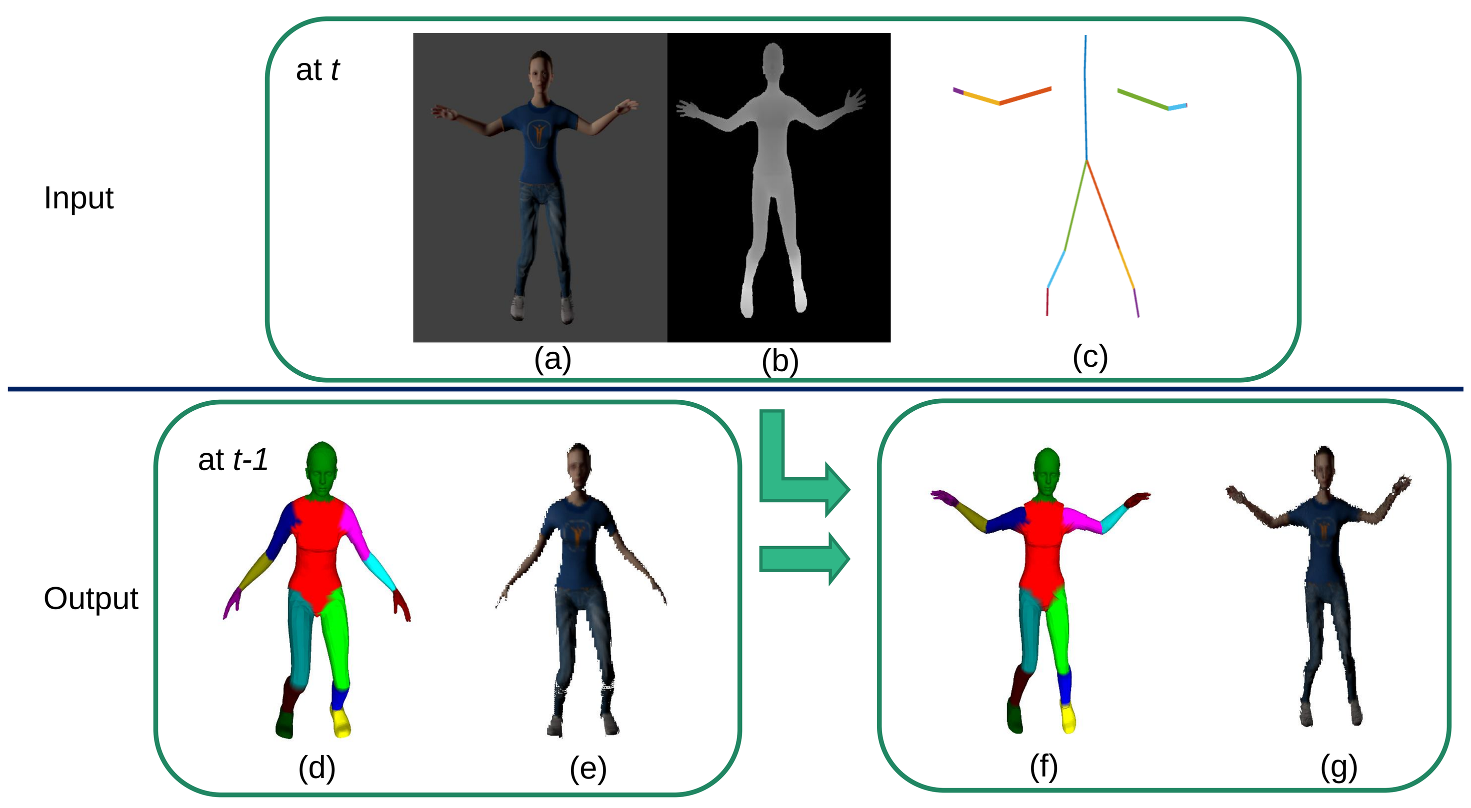
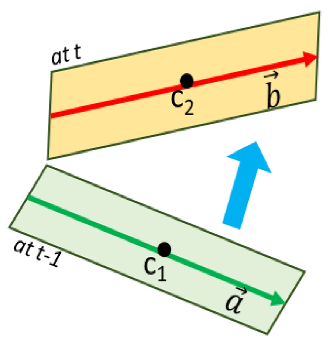


Figure 1: Block diagram of inputs and outputs of our proposed system. (a), (b) and (c) shows colour, depth and skeleton inputs at current frame. (d) and (e) illustrates puppet model and 3D reconstruction outputs at previous frame and finally (f) and (g) are puppet model and 3D reconstruction output at current frame.

Calculation of initial transformation of each body parts



$$v = \vec{a} \times \vec{b} \qquad c = \vec{a} \cdot \vec{b}$$

$$s = \|v\|$$

$$R_{init} = \mathrm{I} + [v]_\times + [v]_\times^2 \frac{1-c}{s^2}$$

$$t_{init} = -R_{init} c_1 + c_2$$

At each frame initial rigid transformation (R_{init}, t_{init}) of each body part is calculated using the angle between the skeleton bones.

Figure 2: Illustration of estimating rigid transformation using the angle between the skeleton bones.

Correspondence Estimation

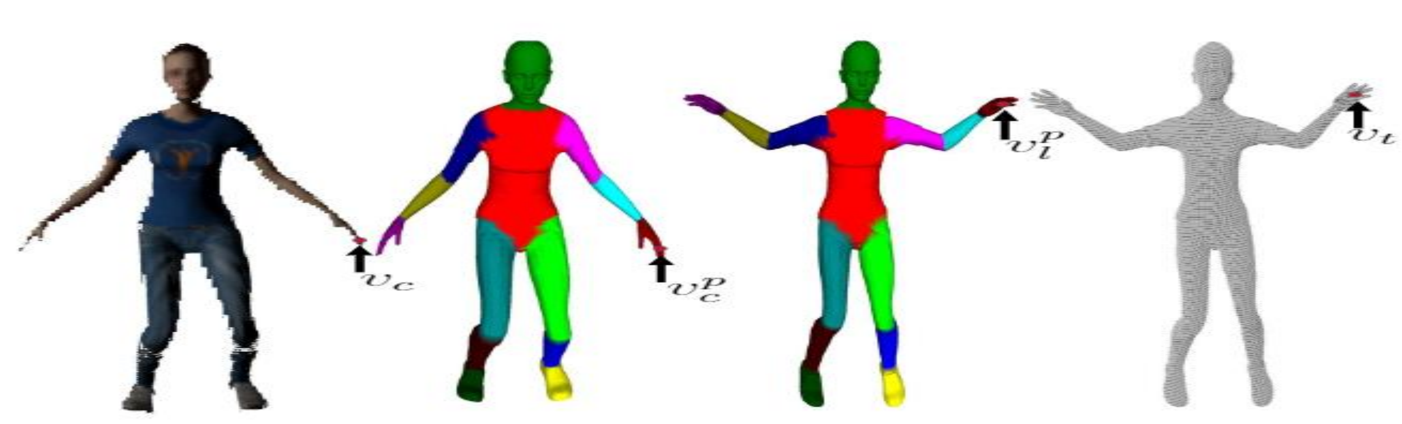


Figure 3: For a point in the reconstruction the nearest neighbour in the puppet model is estimated (shown in first and second images from left to right). The corresponding point in the aligned puppet is used for finding nearest neighbour in the target cloud (shown in third and forth images).

Synthetic Data

- We contribute synthetic data for evaluating non-rigid 3D reconstruction of human subjects .
- The data hashas ground truth for frame-to-frame geometry and skeleton joint detection.
- Publicly available at: https://research.csiro.au/robotics/databases/

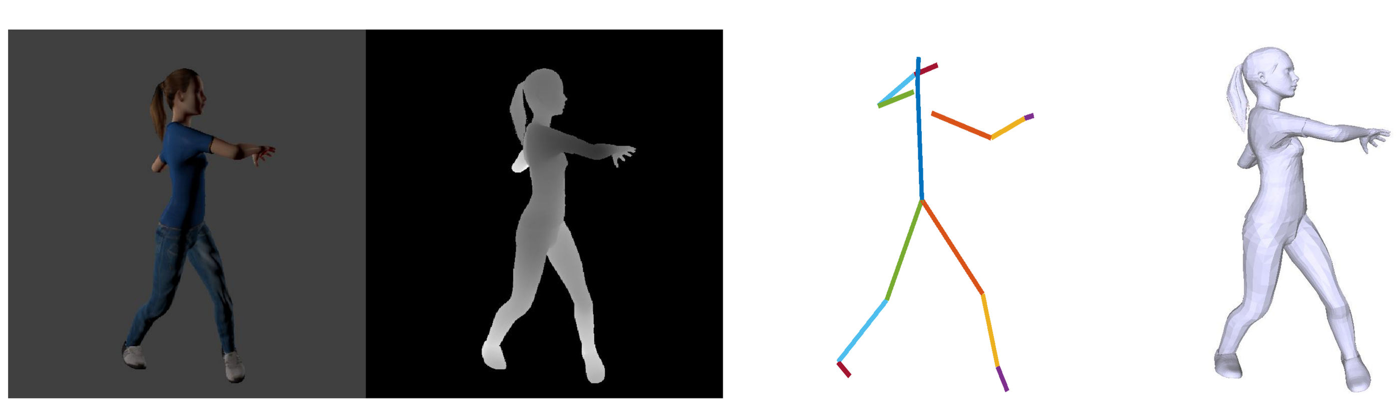


Figure 4: Each frame in the synthetic data consists of RGB image, depth image, skeleton and ground truth geometry (shown from left to right).

Experiments

Advantages of using puppet model's rigid transformation for tracking

	Iteration	Mean	Std	Hausdroff	Outliers
Case1	1	16.1	15.3	0.2155	7599
	2	11.2	11.6	0.2095	6464
	3	9.4	11.0	0.2088	5451
	4	8.6	10.3	0.2070	5027
Case2	1	13.8	13.5	0.2107	7099
	2	10.2	10.6	0.2026	5913
	3	8.5	9.6	0.2013	5179
	4	7.7	8.6	0.1978	4840

Table 1: Qualitative comparison between (case1) tracking without any initialization and (case2) tracking with initialization using puppet model's rigid transformation

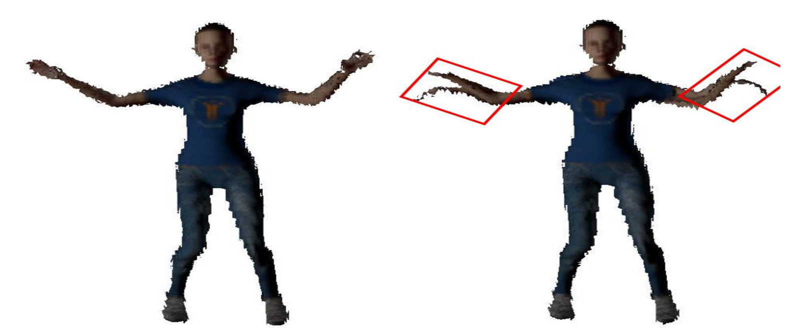


Figure 5: Left with initialization and right without initialization. We can see that in the right image tracking failed and error accumulated in the 3D reconstruction.

Comparison with state-of-the-art approaches

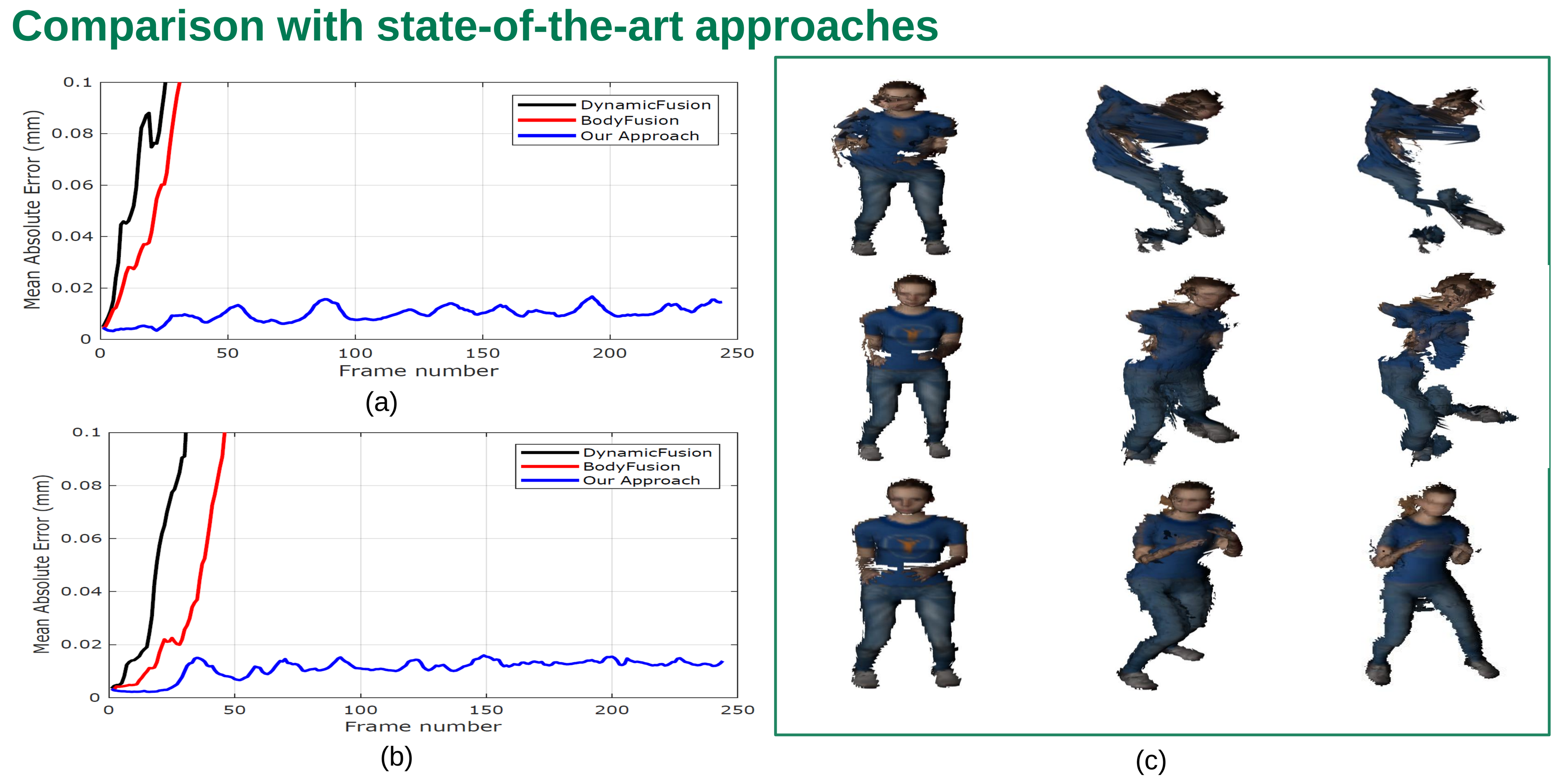
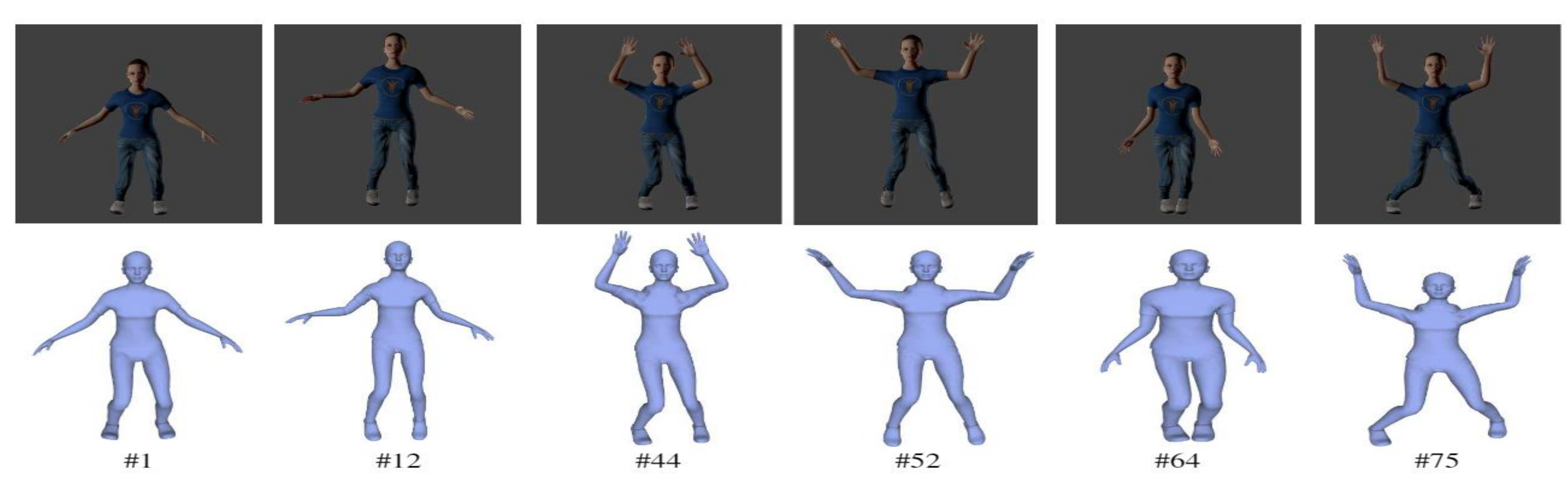


Figure 6: (a,b) - average error on each frame is plotted from the 'Punch Strike' and 'Boxing' data respectively. Black and red correspond to using DynamicFusion[1] and BodyFusion[2] respectively and blue corresponds to our method. (c)- qualitative comparison from 'Boxing' data. The top row shows reconstruction from DynamicFusion[1] and the middle row shows reconstruction from BodyFusion[2] and the bottom row is the reconstruction from our approach.

Qualitative results



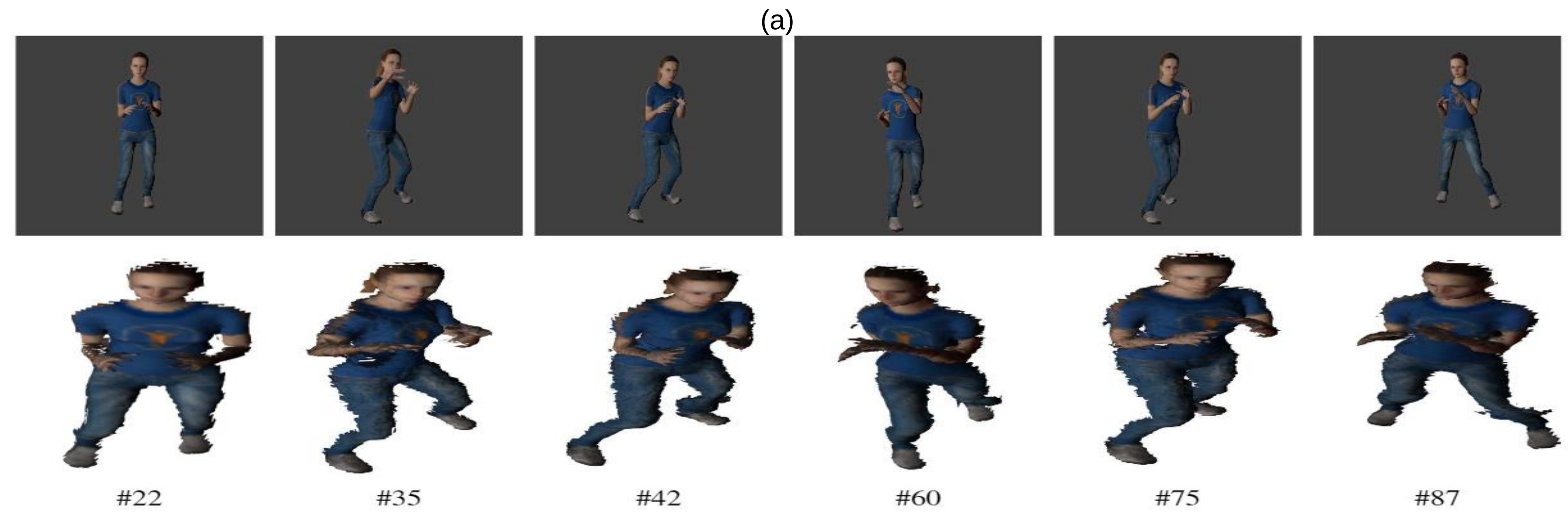


Figure 7: (a)- is motion tracking results from 'Exercise' sequence and (b) is live 3D reconstruction results from the 'Boxing' sequence. For both cases the upper row shows images of different frames and the lower row shows the respective 3D reconstructions. The frame index is shown below.

FOR FURTHER INFORMATION

Shafeeq Elanattil
e Shafeeq.Elanattil@data61.csiro.au

Peyman Moghadam
e Peyman.Moghadam@data61.csiro.au

REFERENCES

[1]: R. A. Newcombe, D. Fox, and S. M. Seitz, "Dynamicfusion: Reconstruction and tracking of non-rigid scenes in real-time," in IEEE Conference on Computer Vision and Pattern Recognition

[2] T. Yu, K. Guo, F. Xu, Y. Dong, Z. Su, J. Zhao, J. Li, Q. Dai, and Y. Liu, "Bodyfusion: Real-time capture of human motion and surface geometry using a single depth camera," in IEEE International Conference on Computer Vision (ICCV), 2017

ACKNOWLEDGEMENTS

The authors gratefully acknowledge funding of the project by the CSIRO and QUT. The institutional support of CSIRO and QUT and the help of several members of staff in the Autonomous Systems Laboratory.