NII International Internship program

SegmentedFusion

Shih Hsuan Yao (Sylvia)

Advisors: Prof. A.Sugimoto

Ass.Prof. D.Thomas

from Aug. 2017 to Feb. 2018

Outline

- Introduction
  - Motivation
  - Related works
  - Contributions
- Method
  - Segmentation
  - Warp field
  - Canonical bounding-box
  - Fusion
- Results
- Conclusions

Presenter: Sylvia
Advisors: Prof. A.Sugimoto, Ass.Prof. D.Thomas

Motivation

- Since depth information has become easily accessible nowadays, many researches have focused on real-time tracking and live scene reconstruction based on a single RGB-D camera.
- By using template as prior assumption, human motion is easy to track but model has limitation to the shape change. On the other hand, template-free reconstruction can be used to estimate non-rigid motion, but still facing challenges on topology change and fast motion.

Related works

- KinectFusion, a system which reconstruction static scene in real-time by using iterative closest point (ICP) to track and using truncated signed distance function (TSDF) to fuse.
- DynamicFusion, estimates the non-rigid motion field and uses efficient volumetric warp to reconstruct the dynamic scene in real-time.
- BodyFusion, combines depth and skeleton information to complete surface reconstruction
- Co-Fusion, segments the scene into different objects while simultaneously tracking and fusing each objects independently.

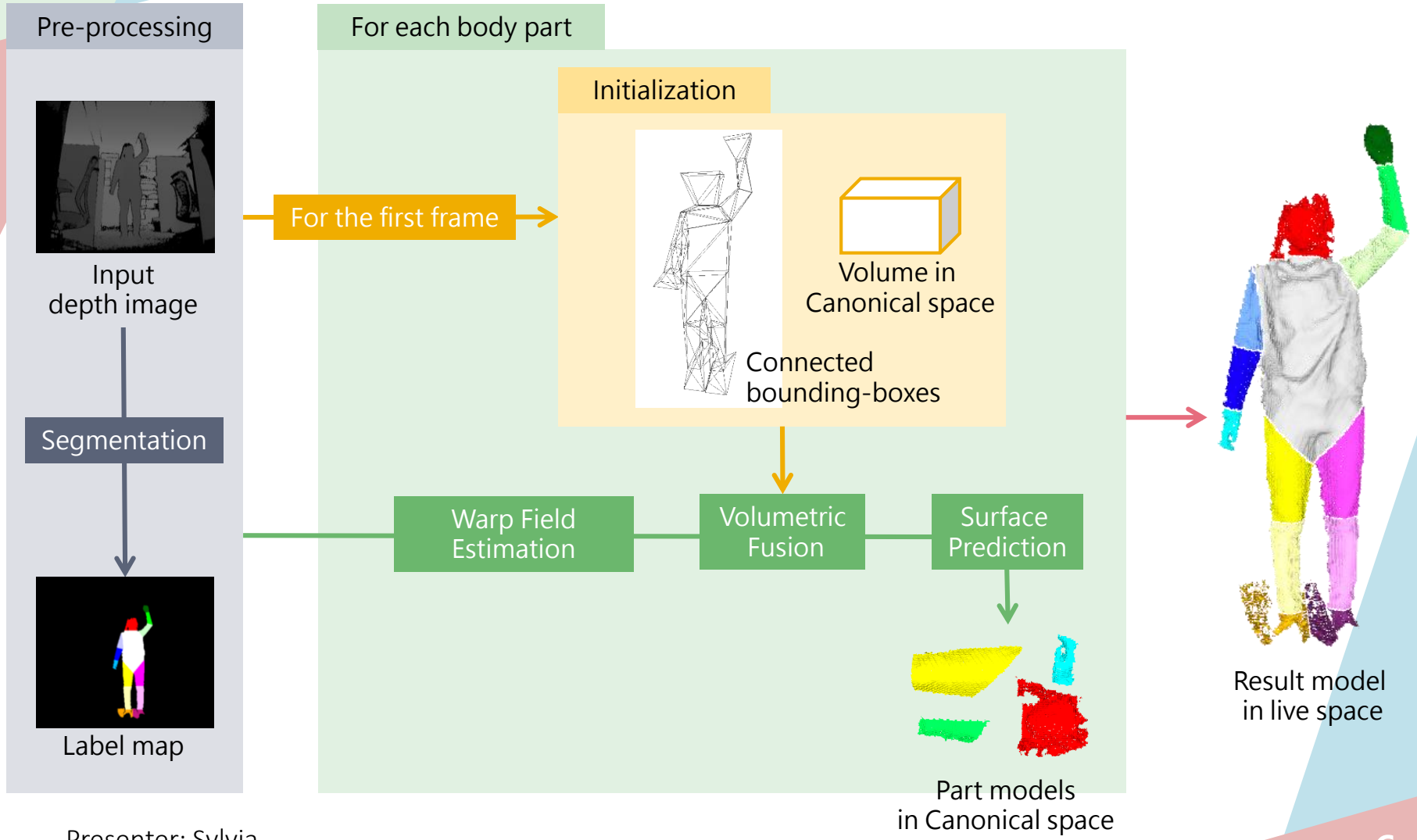
Contributions

- We propose a method which is able to create the canonical bounding-boxes to memory-efficiently scan each body part's model in non-deformed form. By estimating the warp field for each box, the bounding-boxes are deformed and connected to generate a full-body surface.
- Using single depth camera and skeleton, our system captures the motion field, reconstructs 3D human model, and receives good results in handling fast motion and topology change.

Presenter: Sylvia
Advisors: Prof. A.Sugimoto, Ass.Prof. D.Thomas

Method

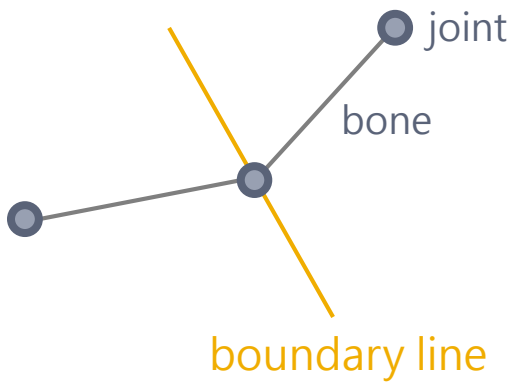
Pipeline



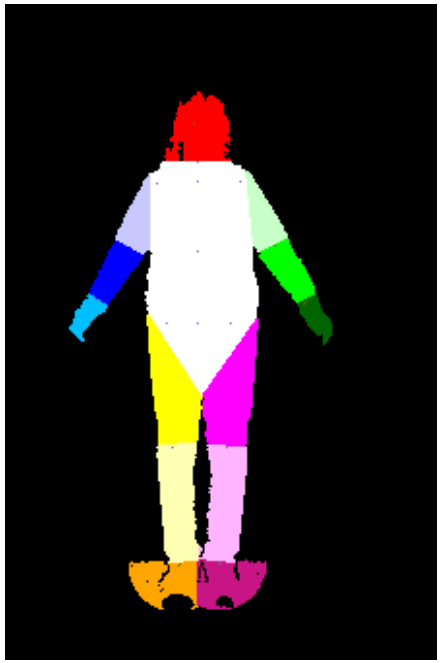
Presenter: Sylvia
Advisors: Prof. A.Sugimoto, Ass.Prof. D.Thomas

Segmentation

- By averaging two adjacent bones' slope, a line equation is calculated, which is perpendicular to the average slope and pass by the intersecting junction.



Line Equation



Label map of Segmentation step

Presenter: Sylvia
Advisors: Prof. A.Sugimoto, Ass.Prof. D.Thomas

Warp field

- The system estimates the deformation field from one frame to another by using skeleton data.
- First, using 3D bones' vectors, we get each body part transform matrix as the following function:

$$\mathbf{R}_t^{t-1} = \mathbf{I} + [v]_\times + [v]_\times^2 \frac{1}{1+c}, \quad \mathbf{t}_t^{t-1} = [B_t(i) - B_{t-1}(i)]^T,$$

$$\mathbf{T} = \begin{bmatrix} \mathbf{R} & \mathbf{t} \\ 0 & 1 \end{bmatrix} \in \mathbb{SE}_3,$$

where $v = B_{t-1}(i) \times B_t(i)$, $c = B_{t-1}(i) \cdot B_t(i)$,
and $[v]_\times$ is the skew-symmetric cross-product matrix

Warp field

- Next, we use dual-quaternion blending to let the surface near the boundary area of two body parts deformed smoothly in space.

$$\mathbf{DLB}(\mathbf{p}) = \mathbf{w}_i(\mathbf{p})\ \mathbf{q}_i + (1 - \mathbf{w}_i(\mathbf{p}))\ \mathbf{q}_{p(i)},$$

- Moreover, ICP is used as refinement step to correct the failed deformation driven by the noise on the data and the twist motion. Singular value decomposition (SVD) is used in implementation while minimizing the distance between reconstructed model and current point cloud.

Presenter: Sylvia
Advisors: Prof. A.Sugimoto, Ass.Prof. D.Thomas

Warp field

- Visual comparisons of the results with and without refinement.

Tracking without refinement

Tracking with refinement

Resource: 031_1027_01_DW_N_R.avi & 031_1027_01_DW_N_N.avi (2fps)

Presenter: Sylvia
Advisors: Prof. A.Sugimoto, Ass.Prof. D.Thomas

Canonical bounding-box

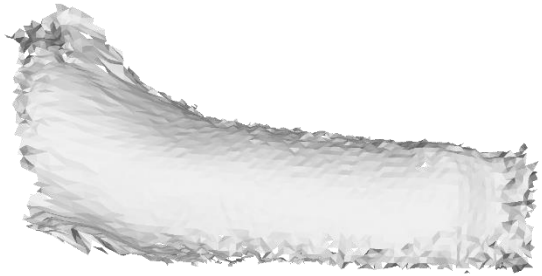
- We define the canonical space for each body part by using the center of mesh as original point and principle vectors as coordinate axes, which are computed by a principle component analysis (PCA).
- The transform matrix of one body part form local to global coordinate is as follow:

$$\mathbf{R} = \left[\mathbf{v1}^T, \mathbf{v2}^T, \mathbf{v3}^T\right], \qquad \mathbf{T}_{loc\to glo} = \begin{bmatrix} \mathbf{R} & \mathbf{t} \\ 0 & 1 \end{bmatrix},$$
$$\mathbf{t} = \mathbf{c}^T,$$

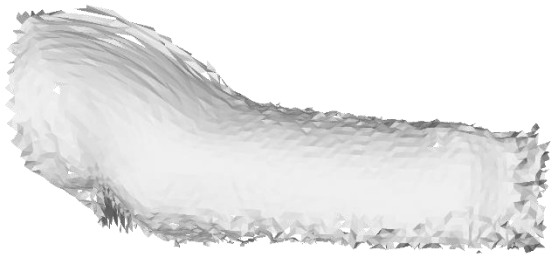
where vector $\mathbf{v1}$, $\mathbf{v2}$, $\mathbf{v3} \in \mathbb{R}^3$ are the principle orthogonal vector, and $\mathbf{c} \in \mathbb{R}^3$ is the center point of the body part

Canonical bounding-box

- In order to make each part stitched properly, we create connected and deformed bounding-boxes in global coordinate. After applying inverse transform matrix, we deform bounding-boxes back into local coordinate, which is used as volume to do fusion.
- Besides, bind matrix of parent joint is used to deform the surface which is influenced by the parent joint' motion. Using volumetric scan fusion, we reconstruct the undeformed surface of each body part.



Arm mesh in bounding-box with undoing parents' motion



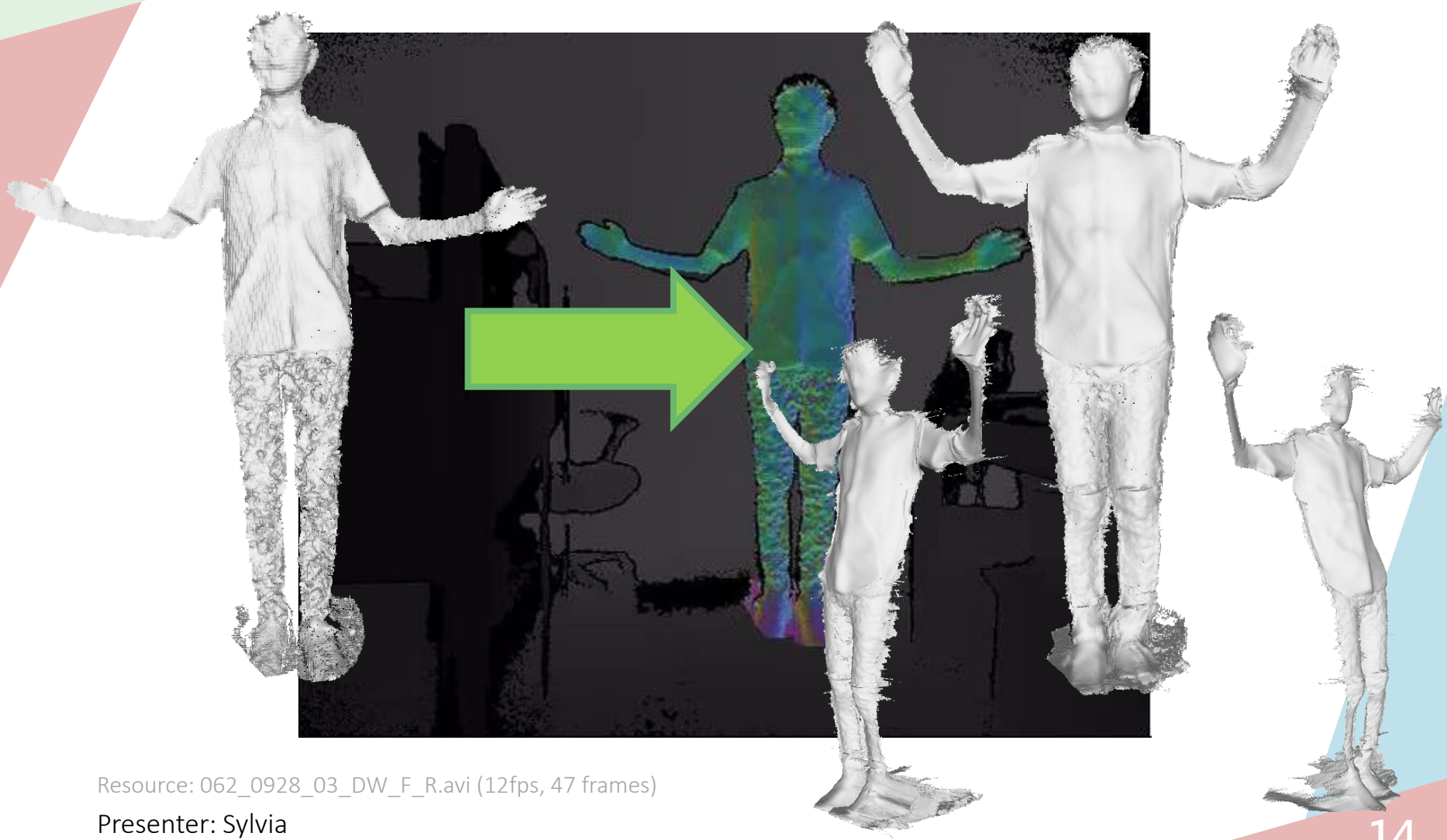
Arm mesh in bounding-box without undoing parent's motion

Presenter: Sylvia
Advisors: Prof. A.Sugimoto, Ass.Prof. D.Thomas

Fusion

- Using model-to-frame warp field, we deform our canonical bounding-boxes to current frame and integrate the live depth value into deformed volume. As DynamicFusion and KinectFusion, TSDF is used to update the signed distances as well as the weights on all voxels.
- Finally, the dense human model represented as triangulation of current frame is rendered by using marching cubes algorithm.

General case



Resource: 062_0928_03_DW_F_R.avi (12fps, 47 frames)

Presenter: Sylvia
Advisors: Prof. A.Sugimoto, Ass.Prof. D.Thomas

Topology change

Resource: 061_0915_01_DW_F_R.avi (12fps, 52 frames)

Presenter: Sylvia
Advisors: Prof. A.Sugimoto, Ass.Prof. D.Thomas

大学共同利用機関法人 情報・システム研究機構
国立情報学研究所
National Institute of Informatics

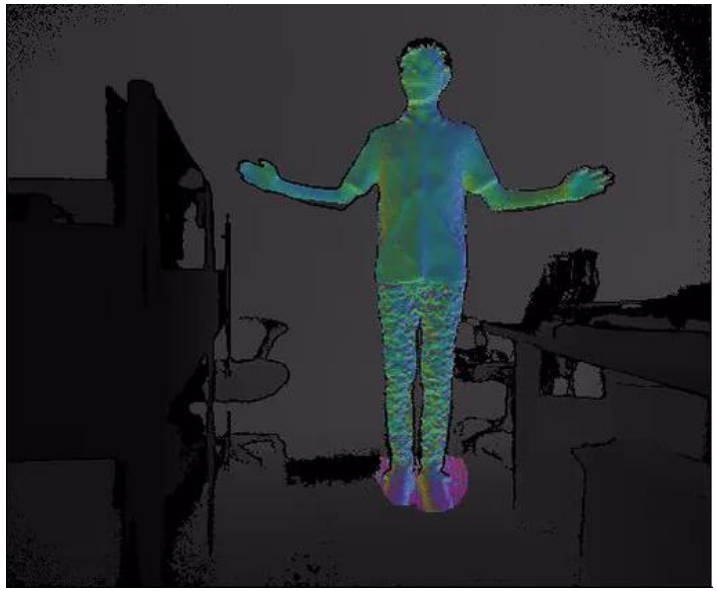
Large frame-to-frame motion

Resource: 041_1027_01_DW_F_R.avi (12fps, 14 frames)

Presenter: Sylvia
Advisors: Prof. A.Sugimoto, Ass.Prof. D.Thomas

Limitations

- Since using ICP as refinement, our system is still limited on the case of fast twist motions. Also, the wrong transformation are found when the number of points on the mesh is not enough.

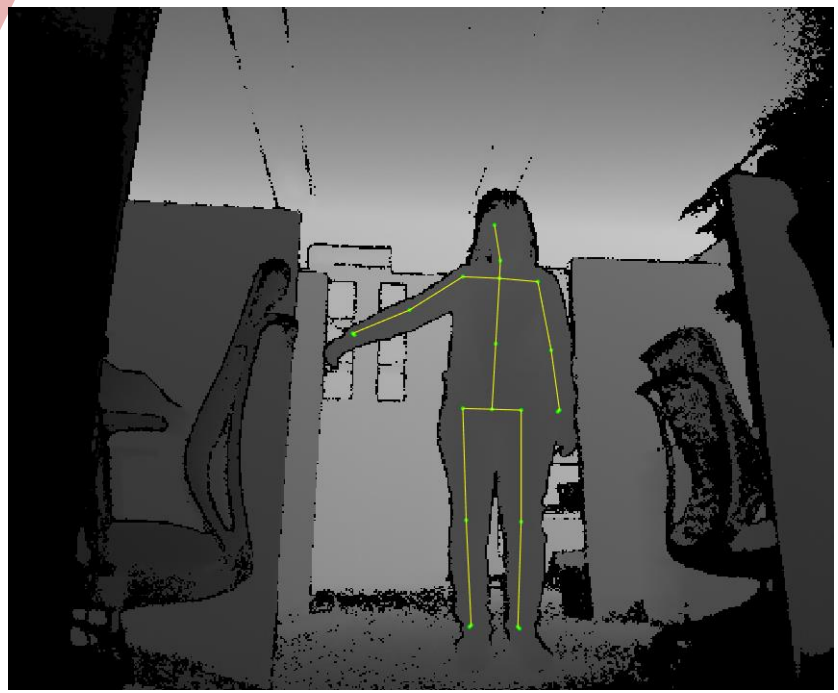


Resource: 062_0928_03_DW_F_R.avi

Presenter: Sylvia
Advisors: Prof. A.Sugimoto, Ass.Prof. D.Thomas

Limitations

- Moreover, the segmentation result has big influence on our system to reconstruct correctly.



input depth frame with skeleton

Label map

Conclusions

- In this work, we propose SegmentedFusion, a system which reconstruct non-rigid human model by using single depth camera with skeleton data.
- Our approach is able to handle the fast motion and topology change by providing a method to segment the body and building the canonical bounding-box for each part, as well as developing a method to estimate a volumetric 6DoF motion field of each part using bone's transform.
- We believe that our system will open the door for the study of 3D scanning with semantic information in general dynamic scene.

Presenter: Sylvia
Advisors: Prof. A.Sugimoto, Ass.Prof. D.Thomas

Reference

- Lorensen, William E., and Harvey E. Cline. "Marching cubes: A high resolution 3D surface construction algorithm." *ACM siggraph computer graphics*. Vol. 21. No. 4. ACM, 1987.
- Besl, Paul J., and Neil D. McKay. "Method for registration of 3-D shapes." *Sensor Fusion IV: Control Paradigms and Data Structures*. Vol. 1611. International Society for Optics and Photonics, 1992.
- Izadi, Shahram, et al. "KinectFusion: real-time 3D reconstruction and interaction using a moving depth camera." *Proceedings of the 24th annual ACM symposium on User interface software and technology*. ACM, 2011.
- Newcombe, Richard A., Dieter Fox, and Steven M. Seitz. "Dynamicfusion: Reconstruction and tracking of non-rigid scenes in real-time." *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2015.
- Yu12, Tao, et al. "BodyFusion: Real-time Capture of Human Motion and Surface Geometry Using a Single Depth Camera."
- Rünz, Martin, and Lourdes Agapito. "Co-fusion: Real-time segmentation, tracking and fusion of multiple objects." *Robotics and Automation (ICRA), 2017 IEEE International Conference on*. IEEE, 2017.
- Dou, Mingsong, et al. "Fusion4d: Real-time performance capture of challenging scenes." *ACM Transactions on Graphics (TOG)*35.4 (2016): 114.

Presenter: Sylvia

Advisors: Prof. A.Sugimoto, Ass.Prof. D.Thomas

Q & A