NII International Internship program

Segmented Fusion

Tracking

20171106

Sylvia

Advisors: Prof. A.Sugimoto

Ass.Prof. D.Thomas

Last Meeting

- Previously
  - Tracking
- To-do
  - New Skeleton [1]
  - Add 3D junction
  - New energy function
  - Co-Fusion [2]

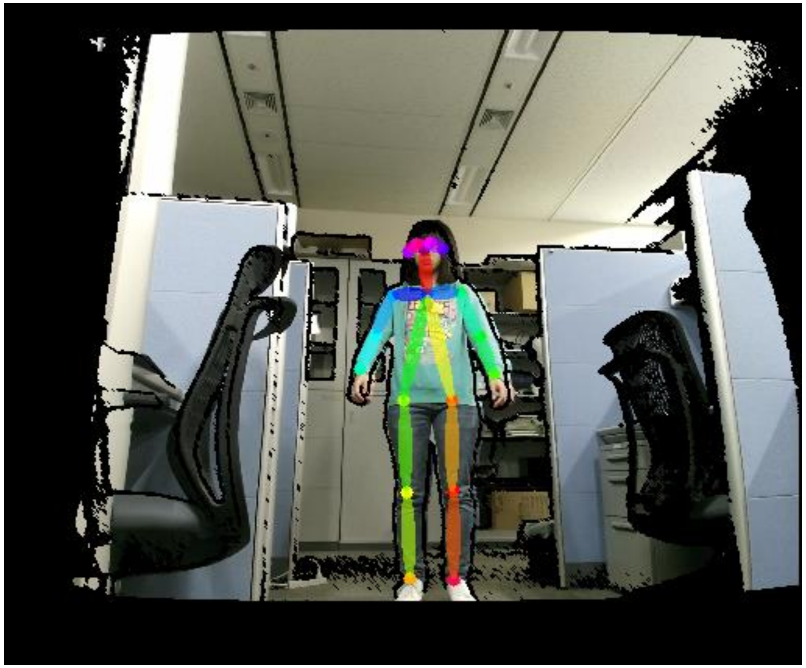
[1] Cao, Zhe, et al. "Realtime multi-person 2d pose estimation using part affinity fields." *arXiv preprint arXiv:1611.08050* (2016).

[2] 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.

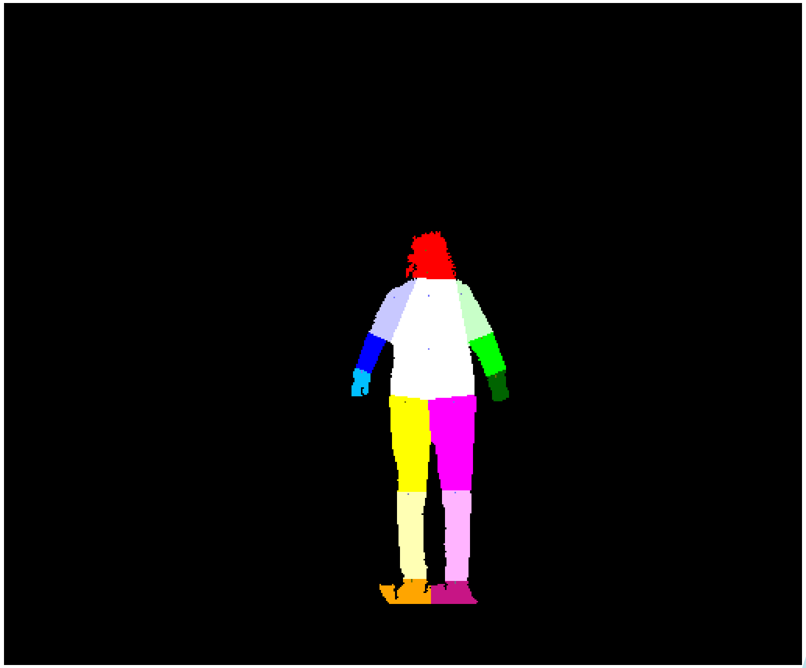
Presenter: Sylvia
Advisors: Prof. A.Sugimoto, Ass.Prof. D.Thomas

New Skeleton

- Created new dataset with depth and color images in the same size and estimated new skeleton by [1]



New skeleton



Segment result

[1] Cao, Zhe, et al. "Realtime multi-person 2d pose estimation using part affinity fields." *arXiv preprint arXiv:1611.08050* (2016).

Presenter: Sylvia

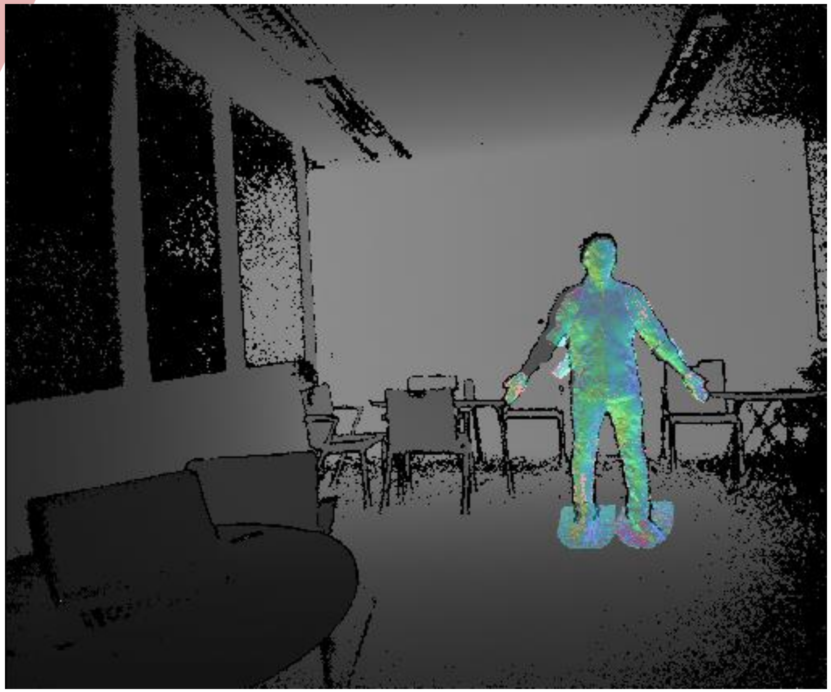
Advisors: Prof. A.Sugimoto, Ass.Prof. D.Thomas

New energy function

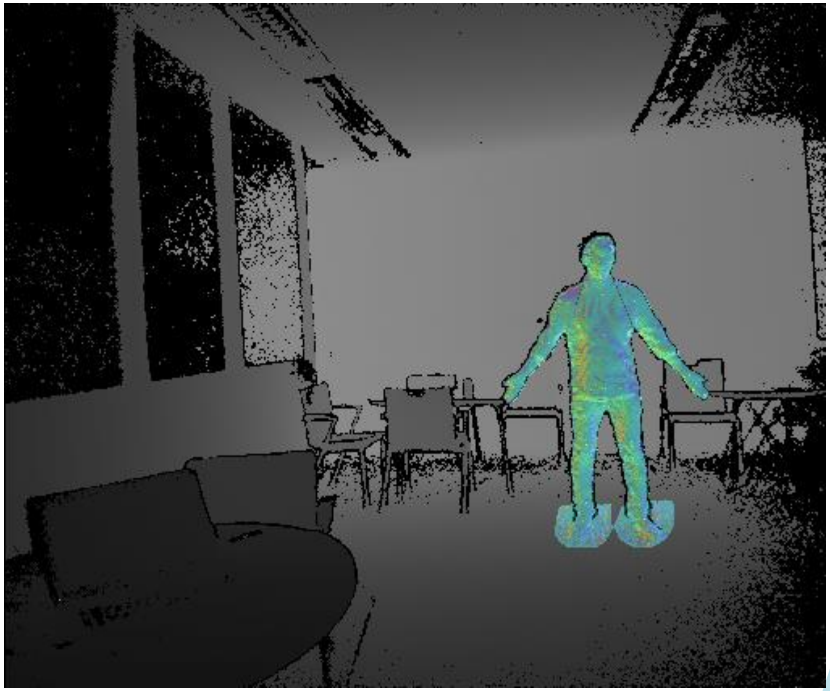
- $E_{data} = \sum_i \sum_j \left\| v_j - c_j \right\|_2$
- I added a constraint that if 2D point projected from c_j is out of the body, set the fixed cost.

- Since it costs time to optimize all transformations at once, I optimize transformations separately. First, I optimize data term independently and get the transformations as initialization. Then, the three term function is minimized.

New energy function



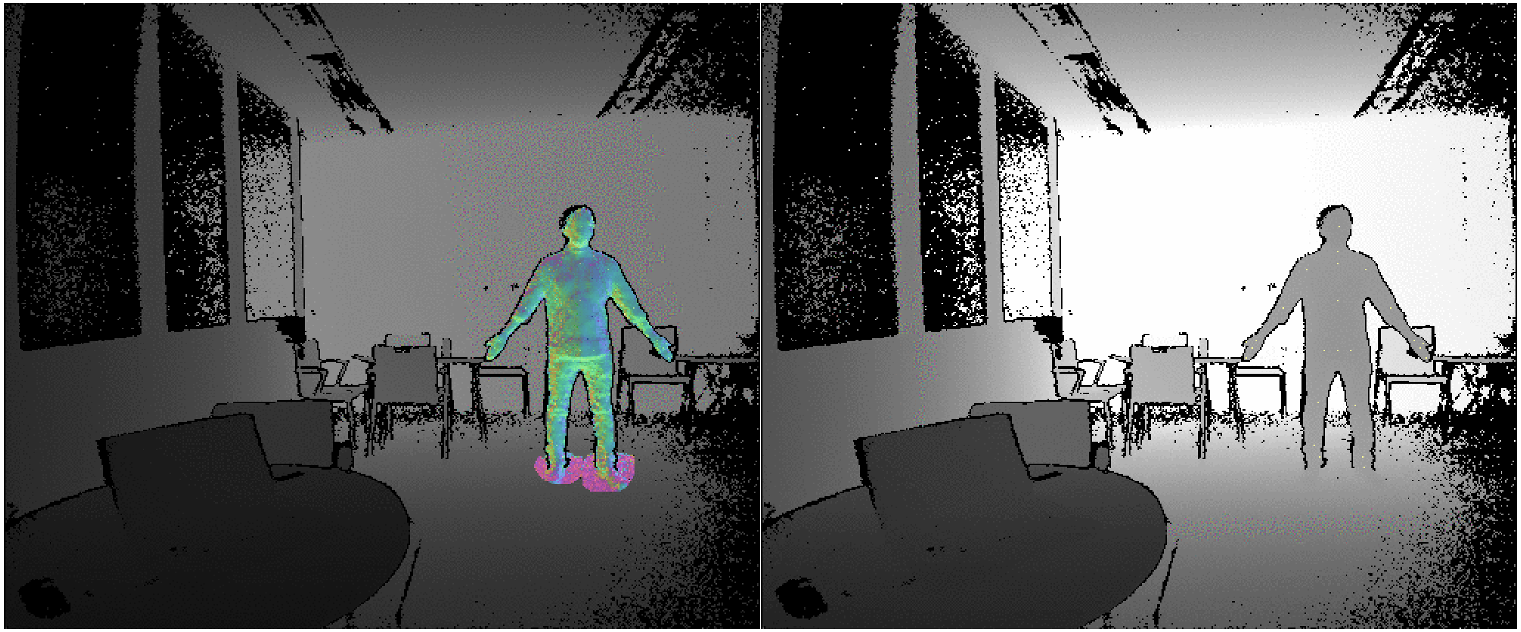
Optimized at once
Time: 331s



Optimized separately
Time: 129s

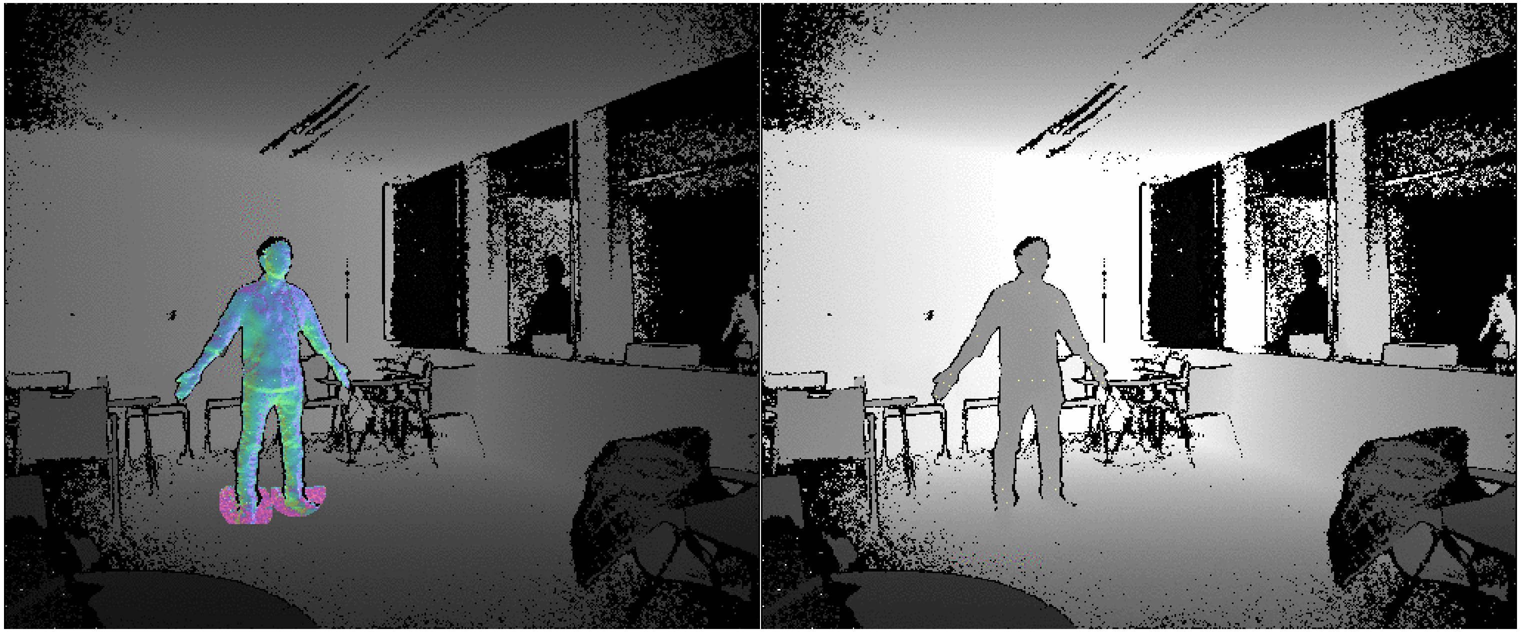
New energy function

- When the camera motion is big:



New energy function

- When the camera motion is big:



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

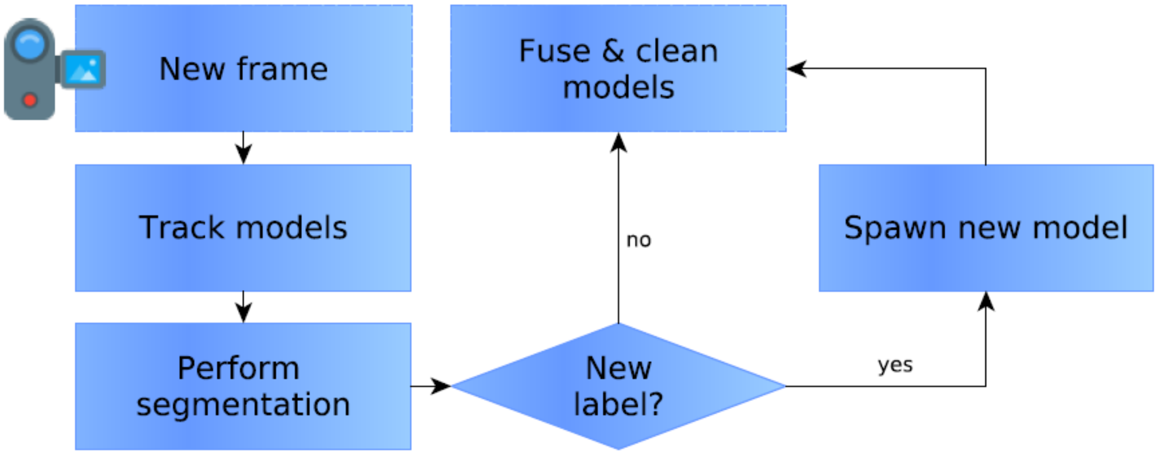
Progress

New energy function

- Result of tracking 20 frames without fusion

System

- Co-Fusion is a dense SLAM system that takes a live stream of RGBD images and segments the scene into different objects while simultaneously tracking and reconstruction their 3D shape in real time.



pipeline

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

Segmentation

- There are two alternative grouping strategies – motion segmentation and object instance segmentation.

Co-Fusion

Motion Segmentation

- After the tracking step, the system have new estimated for the model M_t rigid transformations $\{\mathbf{T_{tm}}\}$.
- The motion segmentation problem is a labeling problem, where the labels are the M_t.
- The cost function has two term: the unary potentials $\varphi_u(x_i)$ and the pairwise potentials $\varphi_p(x_i, x_j)$.

$$E(\underline{\mathbf{x_t}}) = \sum_{i} \varphi_u(x_i) + \sum_{\underline{i<j}} \varphi_p(x_i, x_j)$$

$\mathbf{x_t}$ is labeling result in the frame t

i, j are indices over the super-pixels ranging from 1 to S

Co-Fusion

Motion Segmentation

- **The unary potentials** $\varphi_u(x_i)$ are the estimated ICP alignment costs that apply the rigid transformation associated with each label as defined in E_{icp}^m. For each super-pixel with the outlier label, the cost is determined by the cost of best fitting label.
- **The pairwise potentials** $\varphi_p(x_i, x_j)$ is

$$\varphi_p(x_i, x_j) = \underline{\mu(x_i, x_j)} \sum_{m=1}^{K} w_m \underline{k_m(f_i, f_j)}$$

$\mu(x_i, x_j)$ is the classic Potts models that penalized nearby pixels taking different labels

$k_m(f_i, f_j) = \exp(-\frac{1}{2}(f_i - f_j)^{\mathrm{T}} \Lambda_m (f_i - f_j))$, where f_i f_j are the 6D feature vector and Λ_m is the inverse covariance matrix

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

Motion Segmentation

- After optimize the labeling, the system perform post-processing steps to merge models which have similar rigid transformations, and to spawn the new label(object) when the region of outliers is too large, or to put a missing label out of active model list.

Next step

- Add junction term or as initial