

Overview

Problem. Dense 3D reconstruction of a dynamic foreground subject from a pair of unsynchronized videos with unknown temporal overlap.

Challenges:

1. How to identify temporal overlap in terms of estimated dynamic geometry.
2. How to robustly estimate geometry without knowledge of temporal overlap.

Key Ideas:

1. Define the cardinality of the maximal set of locally rigid feature tracks as a measure of spatio-temporal consistency of a pair of video sub-sequences.
2. Develop a closed-loop track correspondence refinement process to find the maximal set of rigid tracks.

Contributions:

1. We exploit the correlation between temporal alignment errors and geometric estimation errors.
2. We provide a joint solution to the geometry estimation and temporal the video alignment problems.
3. Model-free (i.e. data-driven) framework with wide applicability.

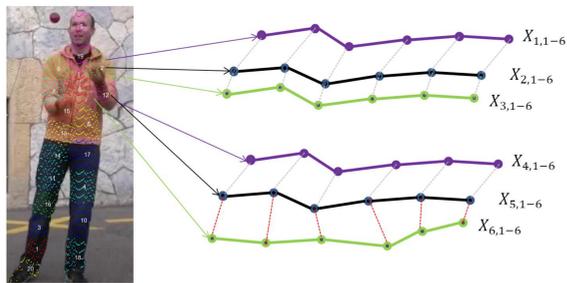
Local Rigidity Test

$$\sum_{i=2}^W \left| \|\mathbf{X}_{m,i-1} - \mathbf{X}_{n,i-1}\|_2 - \|\mathbf{X}_{m,i} - \mathbf{X}_{n,i}\|_2 \right|, \mathbf{T}_n, \mathbf{T}_m \in \mathcal{C}_c$$

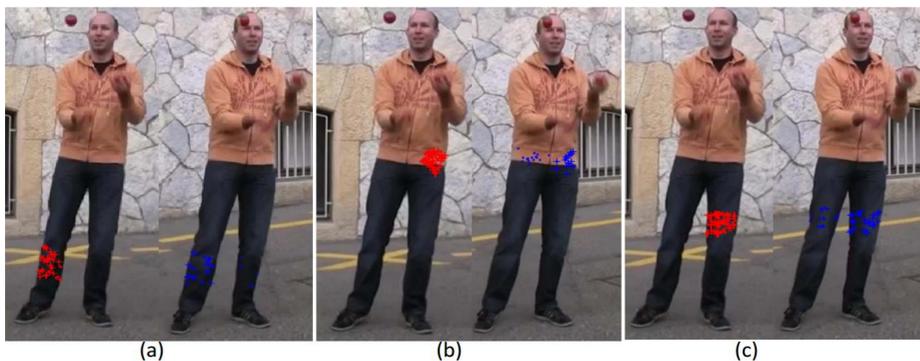
$$\mathbf{T}_i = \{\mathbf{X}_{iw}\}$$

0. The correspondences are computed along the epipolar lines between source and target video frames.

1. If the initial correspondences are within the same local rigid region, distance between 3D points didn't change over a time interval.
2. For correspondences located in different rigid regions, distances between 3D points change over a time interval.
3. For 2D feature points haven't passed rigidity test, we iteratively change it's correspondence (pick other positions on the epipolar line) and do the local rigidity until it pass the test or all candidate positions are tried.



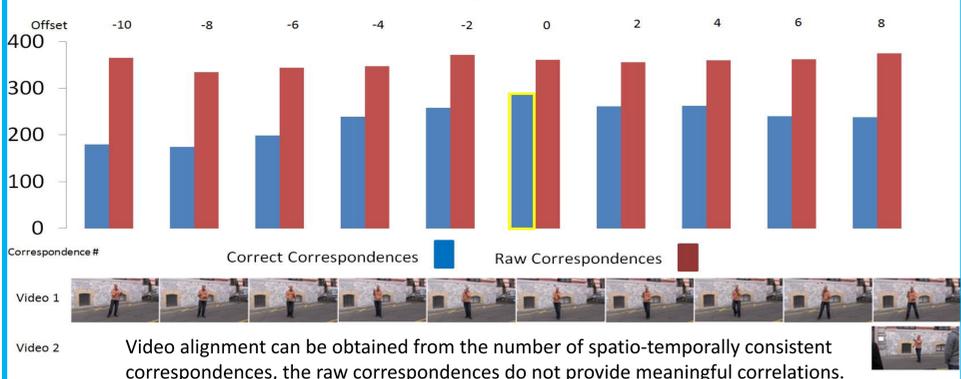
Outlier Detection within Local Rigid Regions



- (a) outliers on the left leg are detected because they located in different rigid parts.
- (b) outliers on the right waist (same rigid part as left waist) are removed because they are far away from majority of the other trajectories
- (c) correct correspondences are the minority (there might be repetitive correspondences in the target frame).

●(+): Feature (inlier) points within one super-pixel in the reference frame.
●(+): Correspondence (inlier) points found in the target frame.

Video alignment

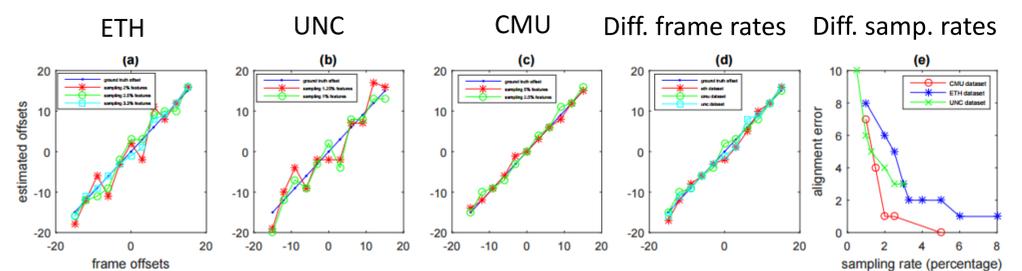


Experiments

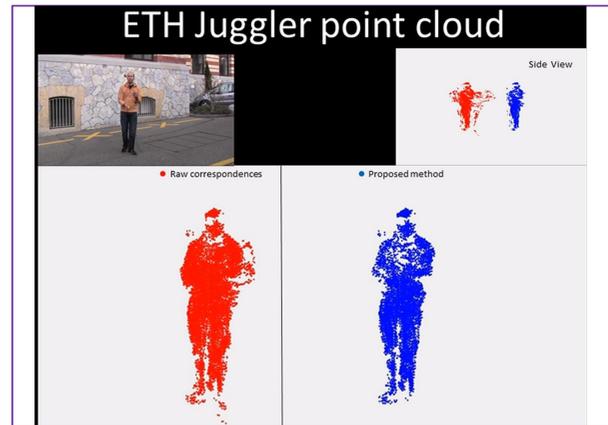
Datasets

Name	# Video frames	GT 3D Points	Synchronized	Moving Cameras	Outdoor Scene
ETH	200	No	Yes	Yes	Yes
CMU	160	Yes	Yes	No	No
UNC	150	No	No	Yes	Yes

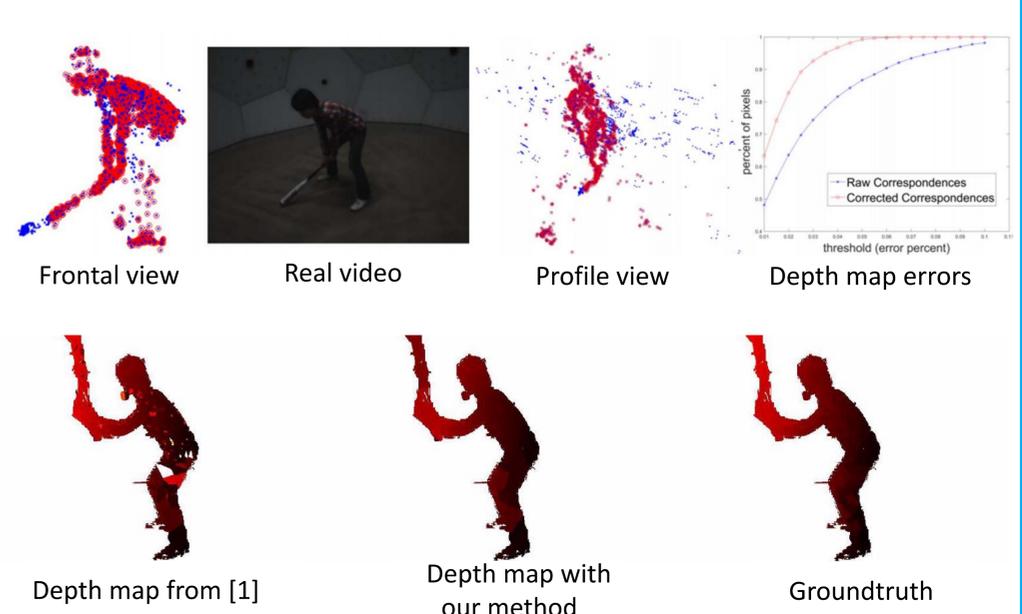
Synchronization Evaluation



Video demo



Results for CMU dataset



Dense reconstruction visualization

