

Multi-Object Shape Estimation & Tracking from Silhouette Cues

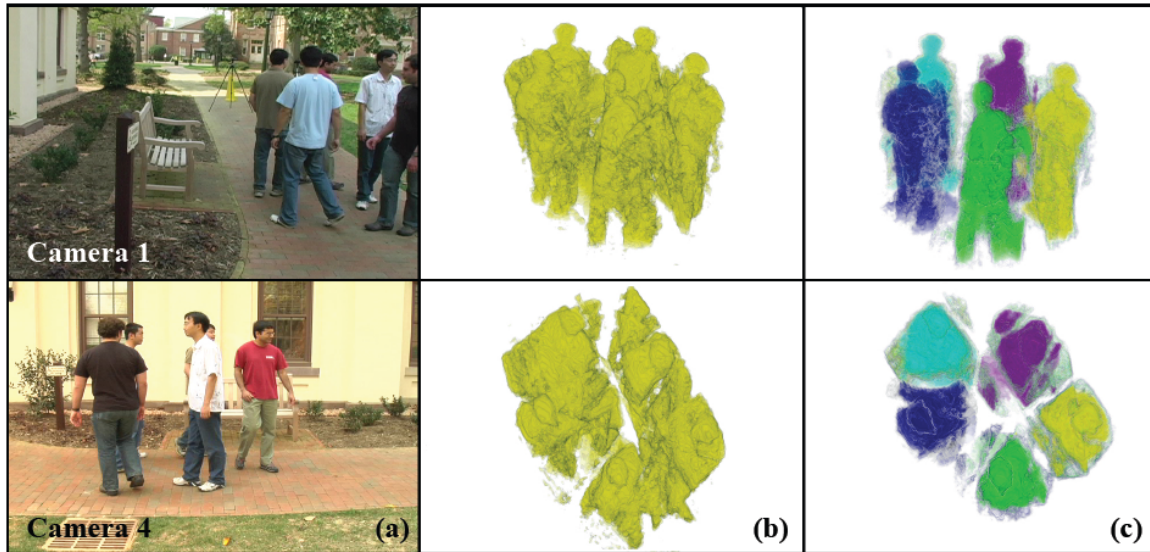
All videos are in .avi format and are coded with DivX 5.2.1 Codec. The newest DivX codec can play them without problem. They can be downloaded for free ([Windows version](#), [Linux version](#), [MAC version](#)). Under Linux, it's probably helpful to install the correct packaged distribution of mplayer.

Four datasets are presented in detail here.

	Cam. No.	Dynamic Obj. No.	Occluder
CLUSTER (outdoor)	8	5	no
LAB (indoor)	15	4	no
BENCH (outdoor)	8	0 - 3	yes
SCULPTURE (outdoor)	9	2	yes

In the videos, the volume transparency and label colors are manually adjusted to give best visualization. The outline around each label is due to volume rendering interpolation artifacts, and is not related to our inference computation.

1. **CLUSTER dataset** ([Click here for video](#))



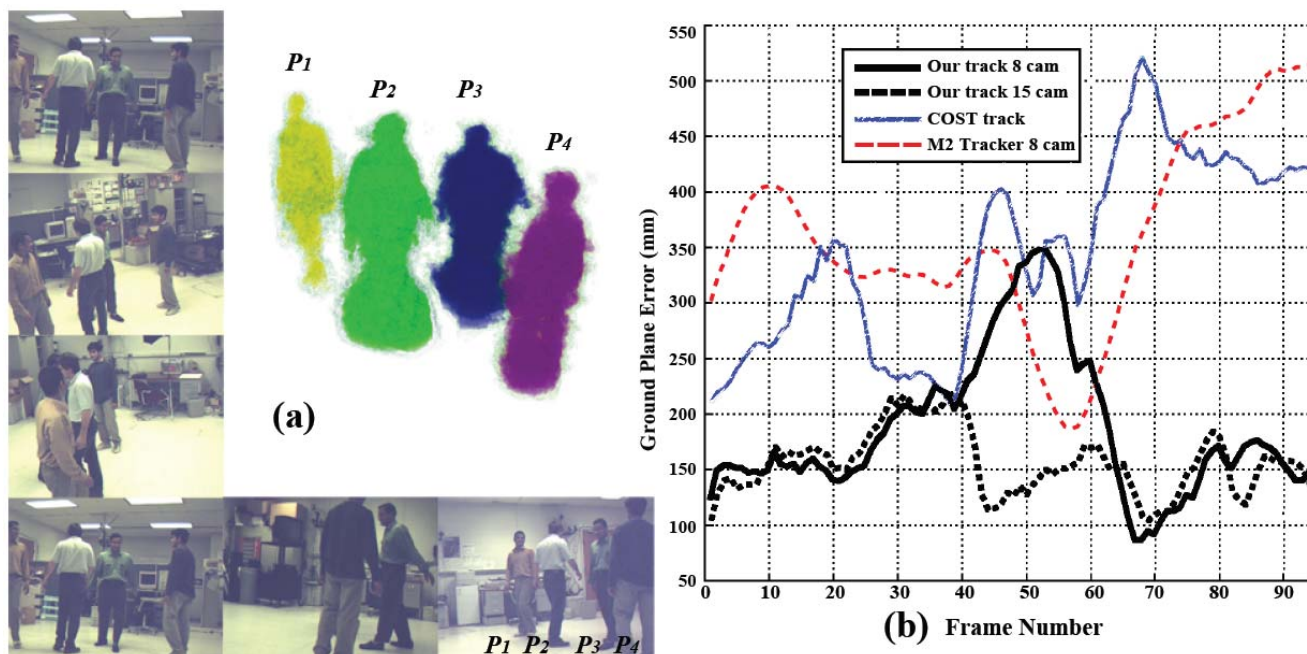
Result from 8-view CLUSTER dataset. (a) Two views at time instance 0. (b) Respective 2-labeled reconstruction. (c) More accurate shape estimation using our algorithm.

We manage to recover individual 3D shapes as well as the tracks accurately, thanks to the multi-label inference and consideration on inter-occlusion relationship between the labels. Comparison with 2-labeled reconstruction (probabilistic visual hull) is given in the video.

Challenges include:

- very clustered outdoor scene
- lighting changes
- shadows
- glass reflections
- color-inconsistency between camera views
- people moving in the background
- similar pants' colors

2. LAB dataset ([Click here for video](#))



Comparison on LAB dataset from Gupta et.al. ICCV 07. (a) 3D reconstruction with 15 views at frame 199 (b) 8-view and 15-view tracking error against the ground truth are compared with results in A. Mittal and L. Davis, IJCV 03 and Gupta et.al. ICCV 07. Mean error on the ground plane estimate in mm is plotted.

The dataset is publicly available. Ground truth tracks are also obtained from Gupta et.al. ICCV 07. This dataset is originally only used for tracking. Now we are able to recover 3D shapes from it. **The plot in the above figure (b) is slightly different from Fig. 5 in the paper:** the 15-view track is also plotted here, which is the case shown in the video (lab.avi). However, even with the exact 8 cameras used in the other two papers, our tracking performance is generally better. The best performance of course is the track with 15-views.

The evaluation in the above figure demonstrates that by recovering the shape, we are able to maintain more accurate tracks in general. At the same time, by having more accurate tracks, we can infer more accurate shapes.

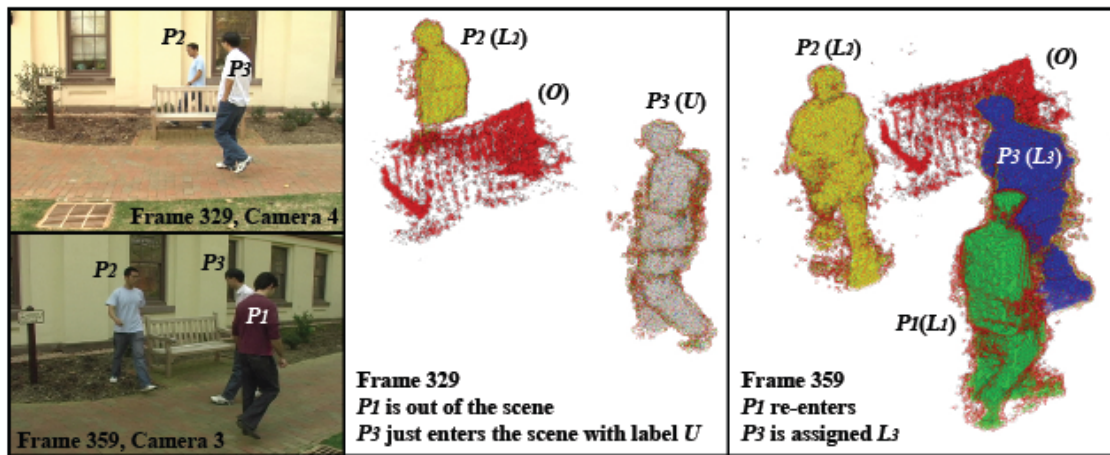
Notice the yellow label (P_1 , in the above figure) partially disappears at one point in the video. It coincides with the part in the above figure (b) where our performance degrades and errors perturb. It's because the person is out of the view in too many of the cameras. And most of his probability is below the threshold for visualization in the video. But the track is still able to be maintained, though with larger error. Our tracking is strongly based on

frame-to-frame localization rather than temporal continuity; the result can expectably be improved by exploring more temporal coherency.

Challenges include:

- poor color contrast
- indoor soft shadows
- clustered scene
- people wear similar color clothes to the background

3. **BENCH dataset** ([Click here for video](#))



During the video, you will see the automatic label updating and maintenance over the sequence even when the dynamic objects leave and re-enter the scene. And the static occluders are recovered simultaneously.

Show the power of the “un-identified” label U for “*dynamic object entering event*” detection and automatic appearance model initialization. It also shows the static occluders can be simultaneously inferred.

Challenges include:

- dynamic objects come in and go out of the scene frequently
- lighting changes
- shadows
- glass reflections
- color-inconsistency between camera views
- people moving in the background
- static occluder

4. **SCULPTURE dataset** ([Click here for video](#))

The dataset is publicly available. Comparison video is also available from Guan et.al. CVPR 2007. The comparison between their video and our video shows the advantages of our algorithm: (1) much more accurate dynamic object shapes can be estimated, which also lead to (2) much more accurate and correct static occluder can be inferred.

Challenges include:

lighting changes
shadows
metal reflections
color-inconsistency between camera views
people moving in the background
static occluder

