

A Framework for Evaluating 6-DOF Object Trackers

Mathieu Garon, Denis Laurendeau and Jean-François Lalonde

Walkman

Kinect

Watering can

Lego





Demos

Working on 6-DOF object pose estimation? Download the dataset here!

http://www.jflalonde.ca/projects/6dofObjectTracking

Context

Evaluating 6-DOF object tracking algorithms properly is key to improve research. However, obtaining the ground truth position of an object w.r.t a camera is a challenging task.

Current datasets for 6-DOF object tracking show serious limitations:

Dataset capture and

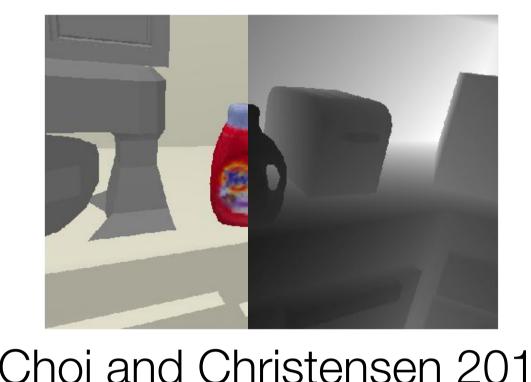
- Unrealistic;
- Small amount of sequences;
- Contain fiducial markers;
- Limited scenarios.

calibration

3mm markers

8 Calibrated Vicon cameras

Kinect V2 (capture sensor)



Choi and Christensen 2013



Our new dataset contains hundreds of realistic sequences, no fiducial markers and a variety of scenarios.

"vcn" reference frame

Object (markers)

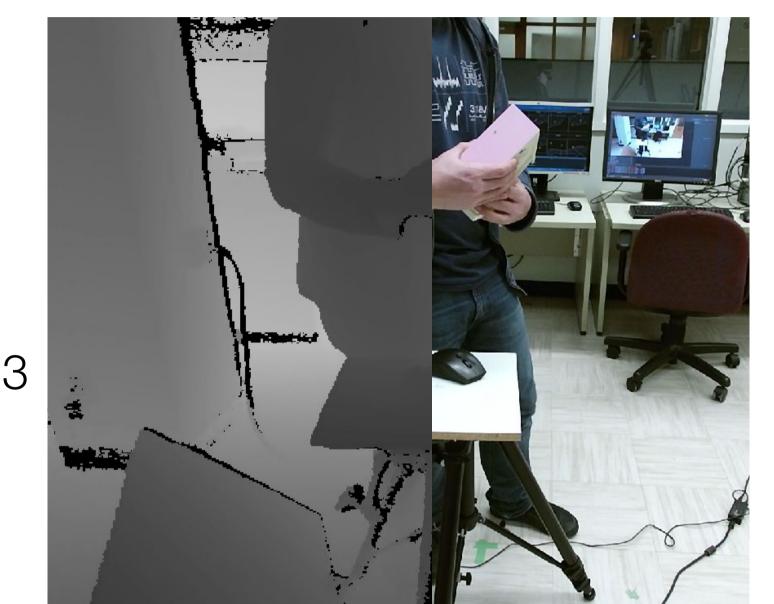
"obim" reference frame

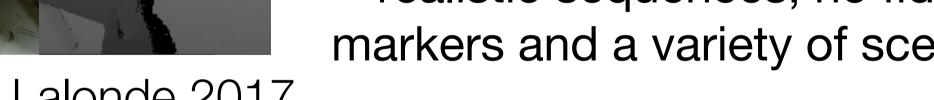
Garon and Lalonde 2017

Kinect (markers)

"kntm" reference frame

kinect pose, ICP is used to fit





Scenarios

Dataset

297 sequences;

62 000 frames;

3 scenarios.

11 objects;

Clock

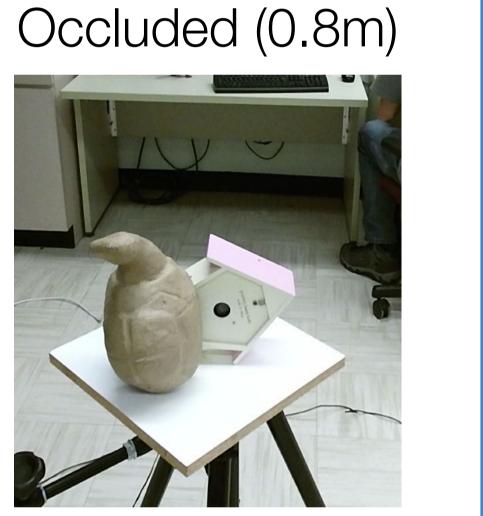
Stability: jittering

- Object is static;
- 3 configurations : near, far, occluded;
- 4 different viewpoints;
- 5 seconds per sequence.

Near (0.8m)



Dragon



Skull

Dog

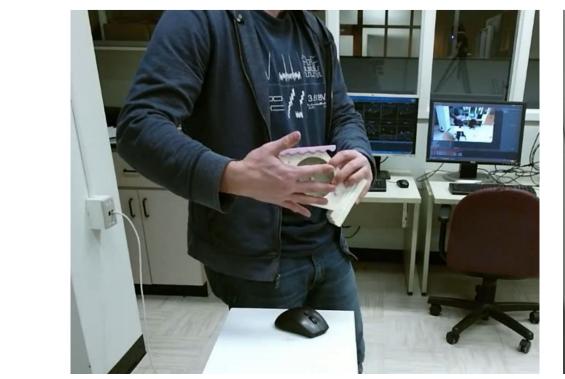
Robustness to occlusion

- Object rotates on a turntable;
- Static occluder from 0% to 75%;
- Horizontal and vertical;
- Occlusion increment of
- 8 seconds per sequence

75% 60% 30%

Free interaction

The object is freely manipulated in 4 different ways: translation only, rotation only, rotation and translation and a hard sequence where multiple occlusions can occur.



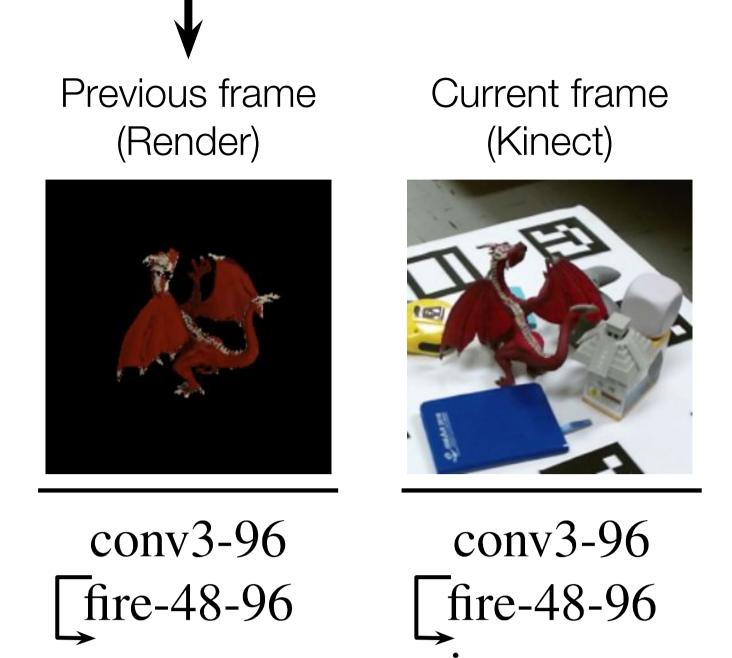


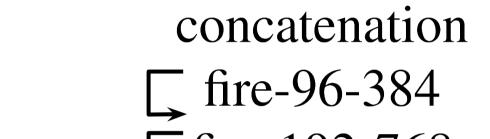


Architecture

Inspired by the architecture of Garon and Lalonde 2017, the previous frame is a render of the object and the network regresses the change in pose.

- Trained on synthetic data only;
- Can generalize to unseen objects;
- More accurate;
- More robust to occlusion;
- More stable;
- Runtime of 6 ms on GTX970M.

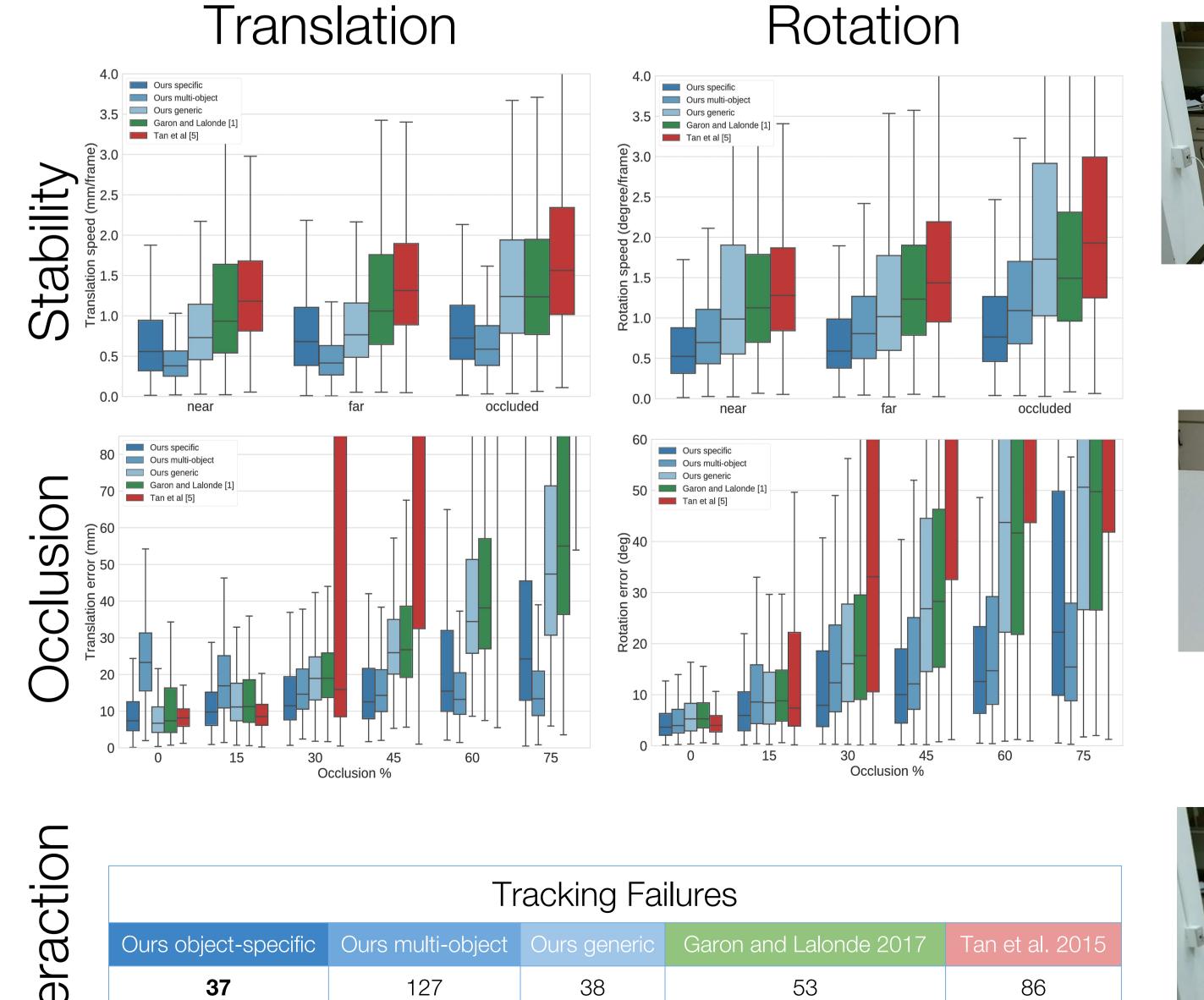




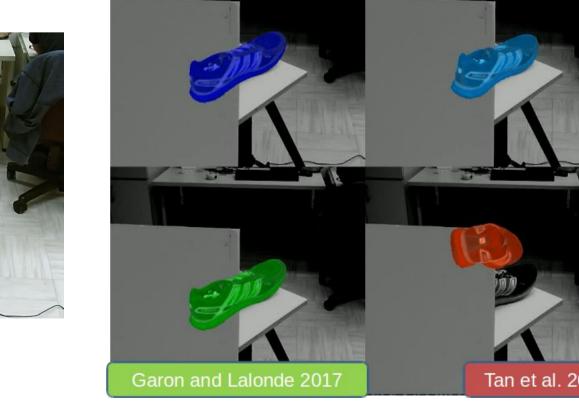
- fire-192-768 fire-384-768 FC-500
- FC-6

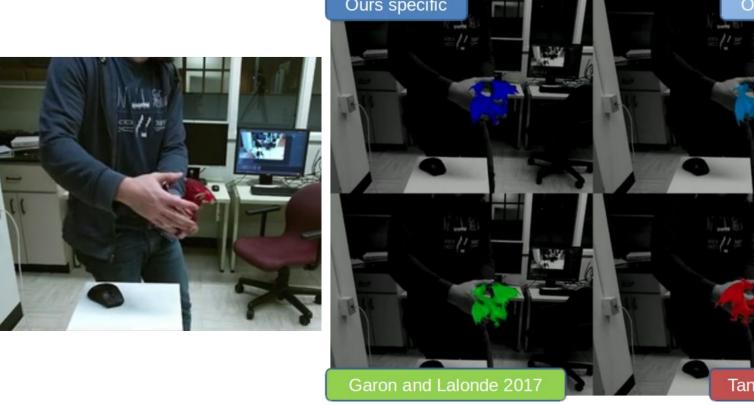
 $\Delta P = [t_x t_y t_z r_x r_y r_z] \longrightarrow P_{t-1} + \Delta P = P_t$

Results









Removing the markers Average of 7 markers

with vicon markers, corresponding

per object; Covers 3.4% of the object pixels per

frame.

find the extrinsic.

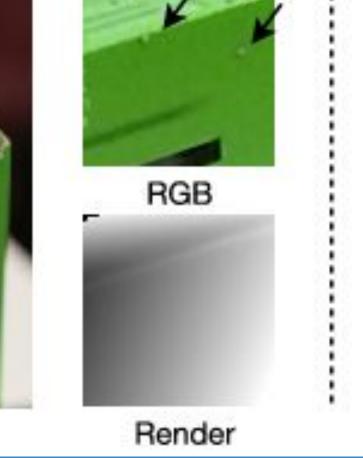


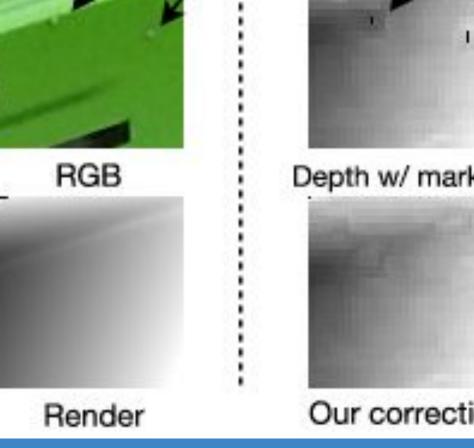
Tknt kinectm: Using a checkerboard pattern Tobj : Using the calibrated

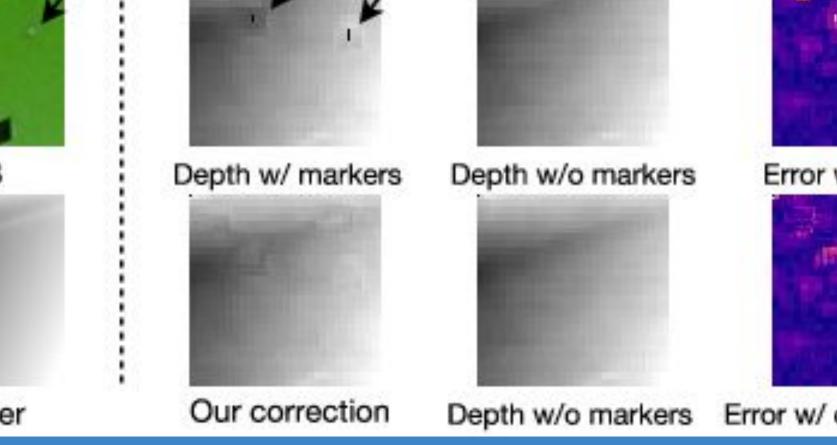
points can be found for the kinect and the mesh with the depth

vicon reference frame. PnP is used to information.









Object (mesh)

Kinect (RGB camera)

"knt" reference frame

