D-PET: A Direct 6 DoF Pose Estimation and Tracking System on Graphics Processing Units

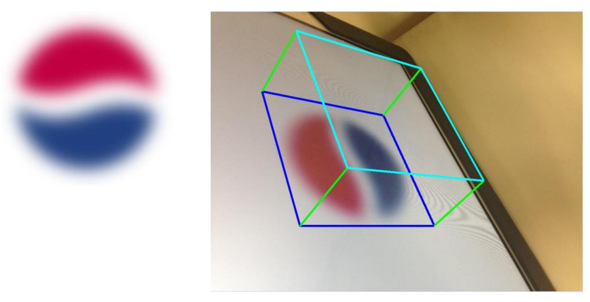
Hung-Yu Tseng, Po-Chen Wu, Yu-Sheng Lin , Shao-Yi Chien

Media IC & System Lab, Graduate Institute of Electronics Engineering, National Taiwan University

Introduction

Real-time recovering 6 degrees of freedom (DoF) object pose is essential in augmented reality (AR) and robotics applications. We built a pose estimation and tracking system that:

- Based on [1] and the proposed 3-scale pose search
- Achieves real-time computation on NVIDIA Jetson TX1
- Able to obtain accurate poses of arbitrary types of target. (not limited to fiducial marker or textured marker)



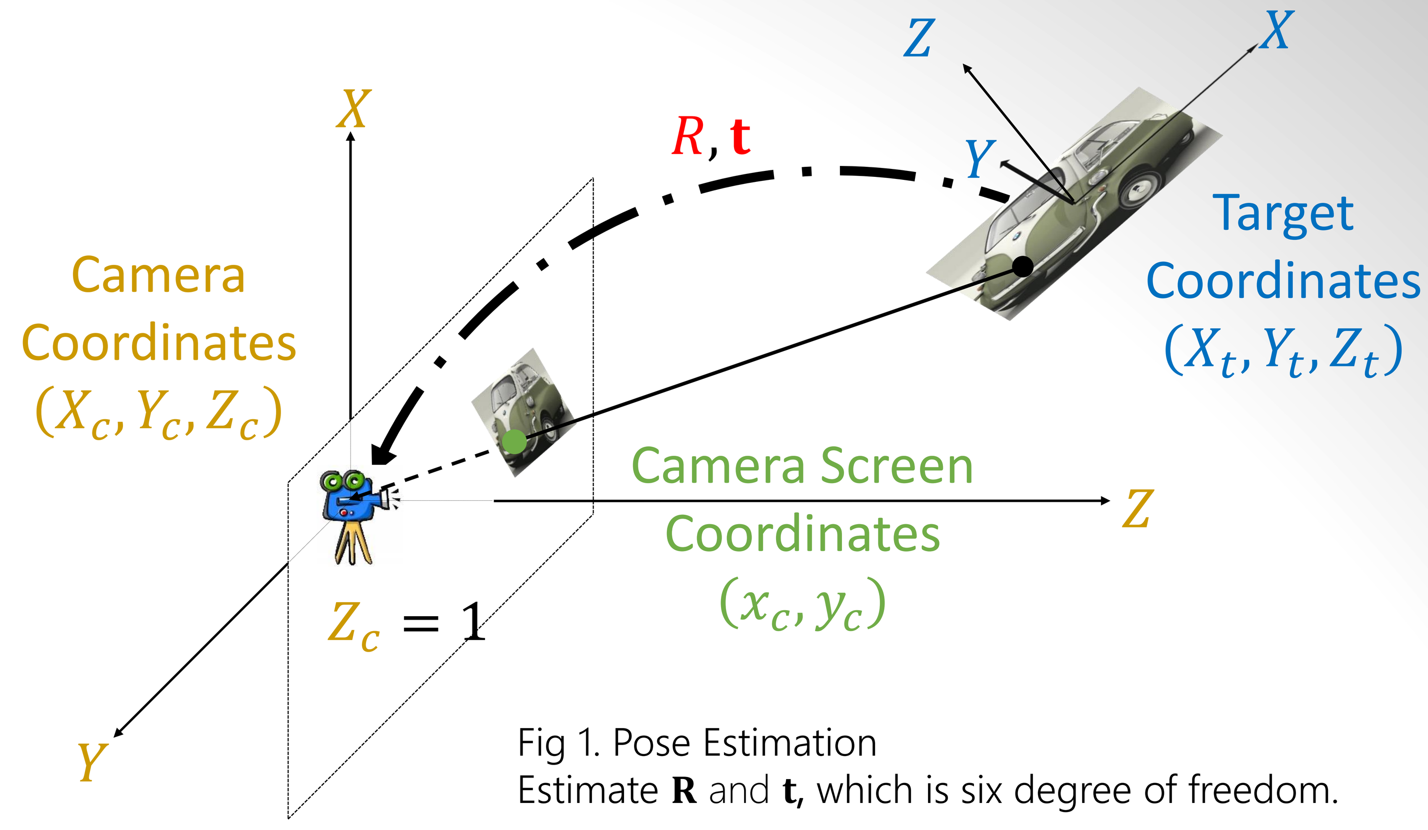


Fig 1. Pose Estimation
Estimate **R** and **t**, which is six degree of freedom.

Designs

- The system aims to find the pose with minimum appearance distance

$$E_a(\mathbf{p}) = \frac{1}{n_t}\sum_{i=1}^{n_t}|I_c(\mathbf{u}_i) - I_t(\mathbf{x}_i)| \qquad \mathbf{x} = (X_t, Y_t) \quad \mathbf{u} = (x_c, y_c) \qquad (1)$$

- Pose Estimation Unit
  - Computes the initial pose
  - Implemented based on the approximated pose estimation (APE) scheme
  - Processes each of the poses parallely
  - Special memory allocation accelerates the parallel computation
    - 4-2 floats memory allocation for storing pose parameters
    - Camera image stored in texture memory
    - Accesses target information via load uniform instructions (LDU)
- Pose Tracker
  - Realizes accurate pose tracking
  - Applies a 6D pose search pattern to perform 3-scale pose search
    - 5 points in each of the rotation dimensions, 7 points in each of the translation dimensions , totally 42875 points for the search pattern

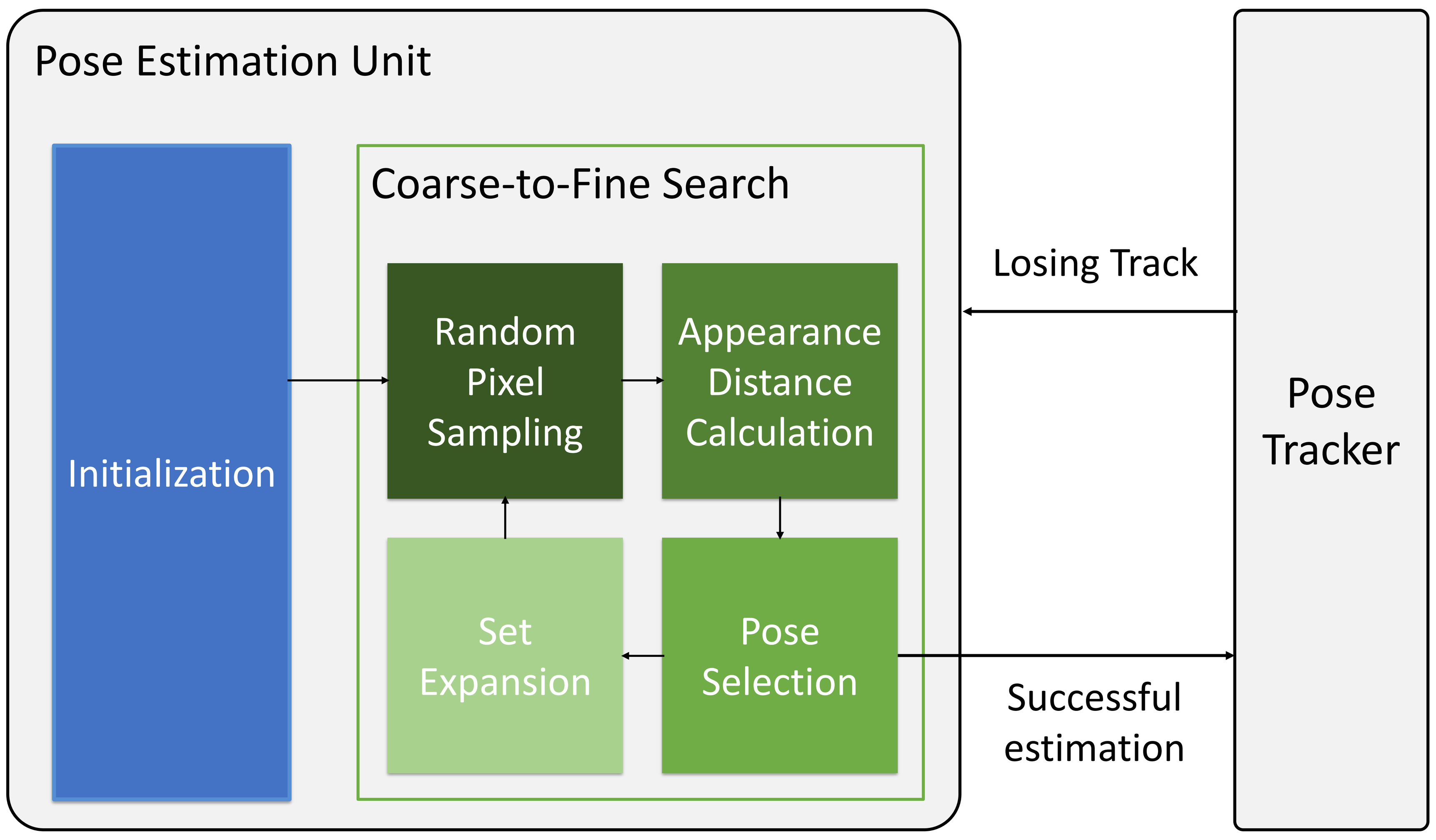


Fig. 2 Proposed D-PET system

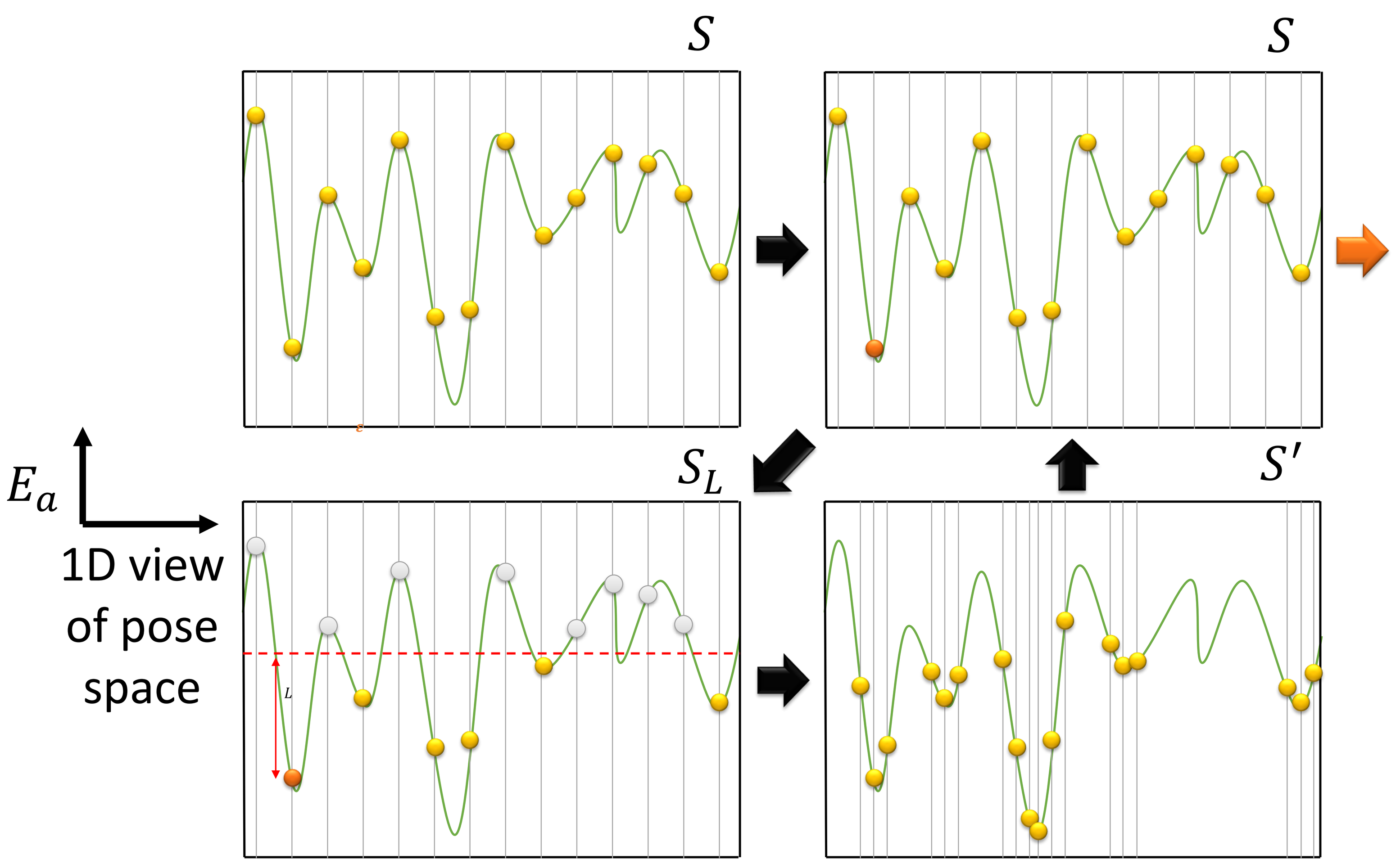


Fig. 3 Approximated Pose Estimation

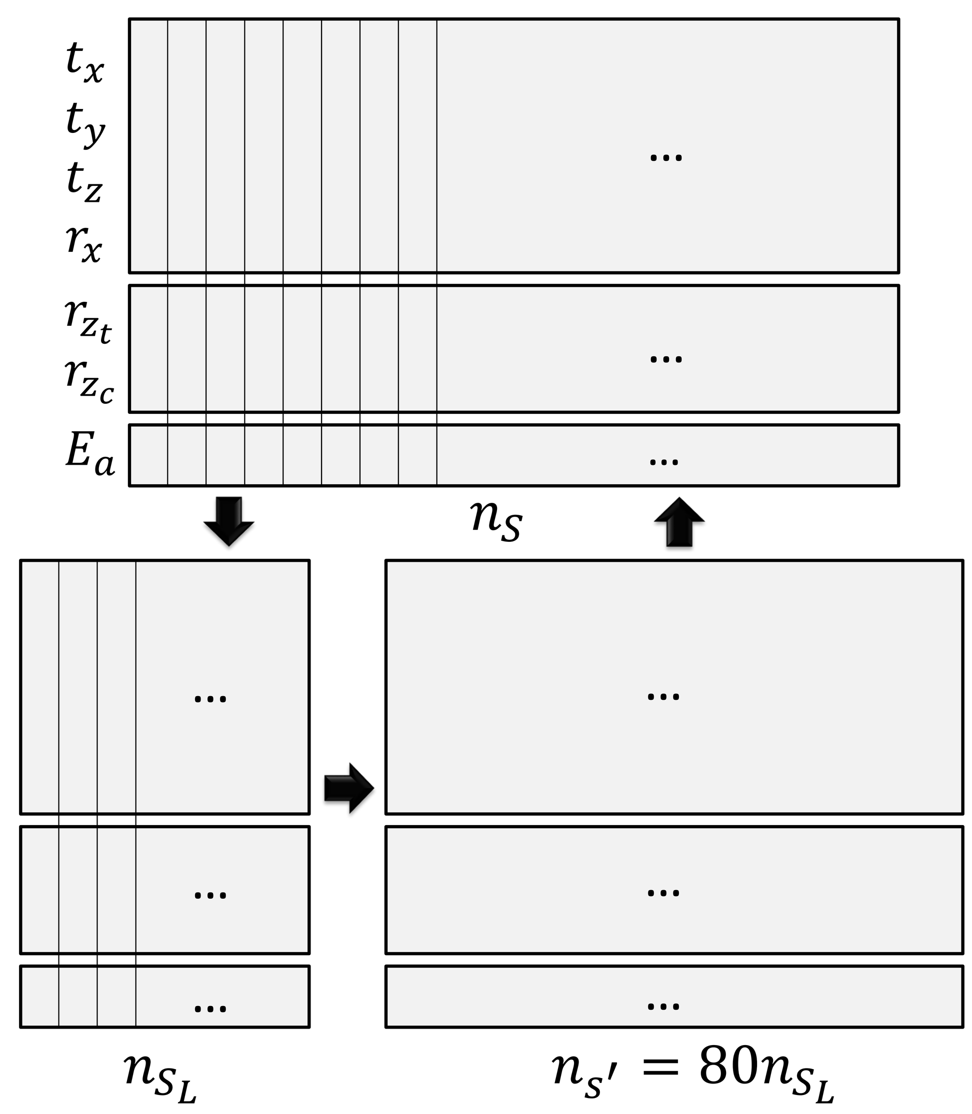


Fig. 4 Memory allocation for pose parameters and appearance distances

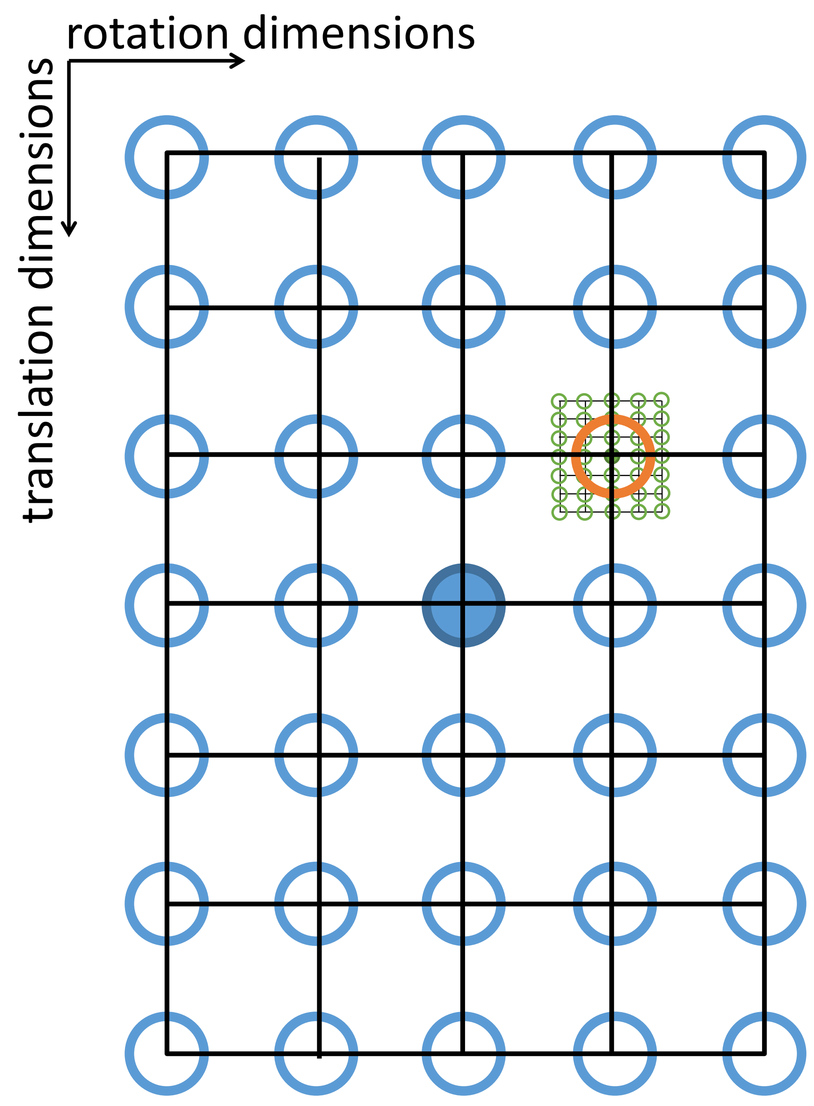
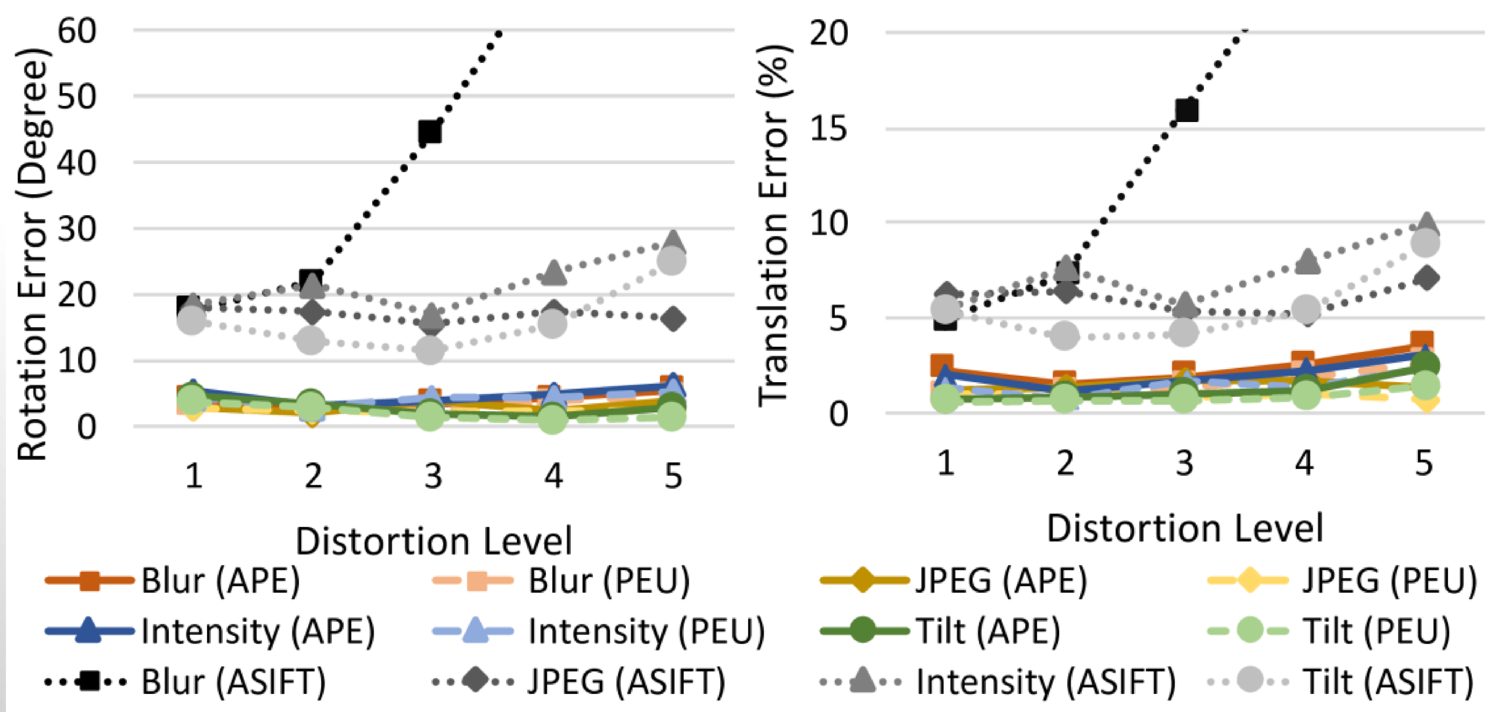


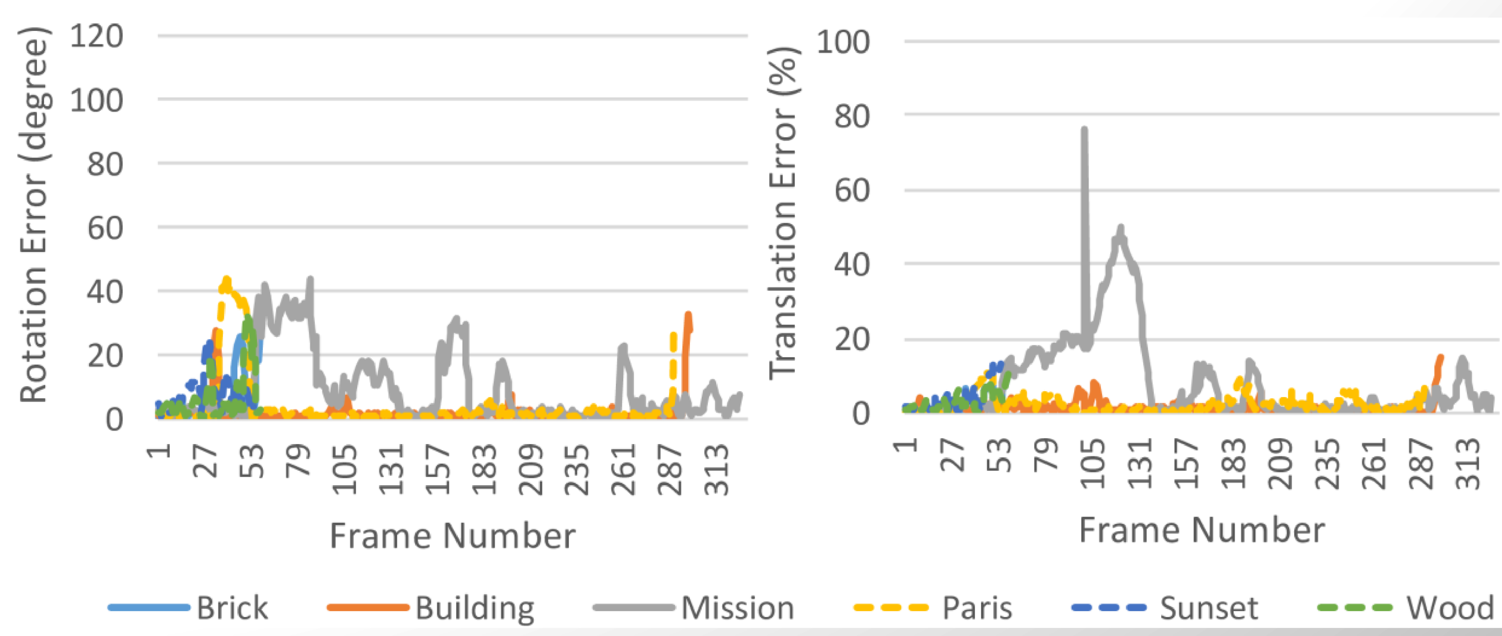
Fig. 5 Proposed 6D pose search pattern

Results

- Evaluation
  - Rotation error (degree) : $E_{\mathbf{R}} = acsd\left(\left(Tr(\mathbf{R}^{\mathrm{T}}\cdot\hat{\mathbf{R}}) - 1\right)/2\right)$
  - Translation error (%) : $E_{\mathbf{t}} = \|\hat{\mathbf{t}} - \mathbf{t}\| / \|\hat{\mathbf{t}}\| \times 100$
- Compare Pose Estimation Unit with APE [1] and ASIFT using the synthetic dataset [1]
- Evaluate Pose Tracker using the real dataset [2]

System	Core Method	Platform	FPS	Resolution
Proposed	Direct Method	NVIDIA Jetson TX1	11	640 × 360
Schaeferling [3]	SURF	2 Xilinx Spartan-3E 2 ARM Cortex-A9	0.94	640 × 480
Rister [4]	SIFT	NVIDIA Tegra 250	7.5	320 × 240
Wang [5]	SIFT	Adreno320	5.9	320 × 256





Conclusion

https://goo.gl/9CCfcV

We propose D-PET, a direct 6 DoF Pose Estimation and Tracking system implemented on an embedded GPU. Compared to the state-of-the-art feature based systems, it is able to deal with more general targets and performs favorably in terms of accuracy and robustness.

Reference:

[1] H.-Y. Tseng, P.-C. Wu, M.-H. Yang, and S.-Y. Chien, “Direct 3D pose estimation of a planar target,” WACV, 2016.
[2] S. Gauglitz, T. H¨ollerer, and M. Turk, “Evaluation of interest point detectors and feature descriptors for visual tracking,” IJCV, 2011.
[3] M. Schaeferling, U. Hornung, and G. Kiefer, “Object recognition and pose estimation on embedded hardware: Surf-based system designs accelerated by fpga logic,” IJRC, 2012.
[4] B. Rister, G. Wang, M. Wu, and J. R. Cavallaro, “A fast and efficient sift detector using the mobile gpu,” ICASSP, 2013.
[5] G. Wang, B. Rister, and J. R. Cavallaro, “Workload analysis and efficient opencl-based implementation of sift algorithm on a smartphone,” GlobalSIP, 2013.