DodecaPen: Accurate 6DoF Tracking of a Passive Stylus

Po-Chen Wu[*†], Robert Wang[†], Kenrick Kin[†], Christopher Twigg[†], Shangchen Han[†]
Ming-Hsuan Yang[‡], Shao-Yi Chien[*]

[*]National Taiwan University, [†]Oculus Research, [‡]UC Merced

Outline

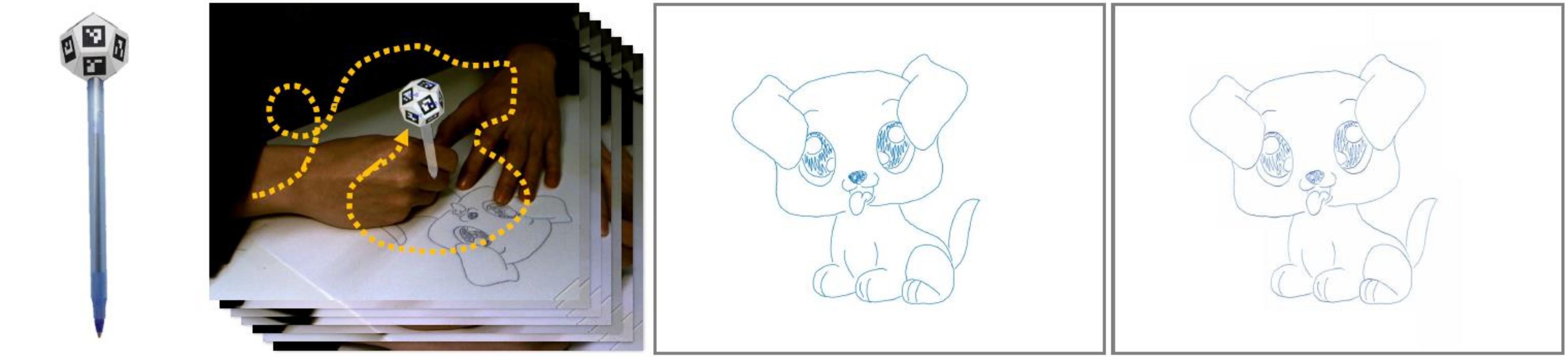
- Introduction
- Proposed System
- Experimental Results
- Conclusion

Outline

- Introduction
- Proposed System
- Experimental Results
- Conclusion

Introduction

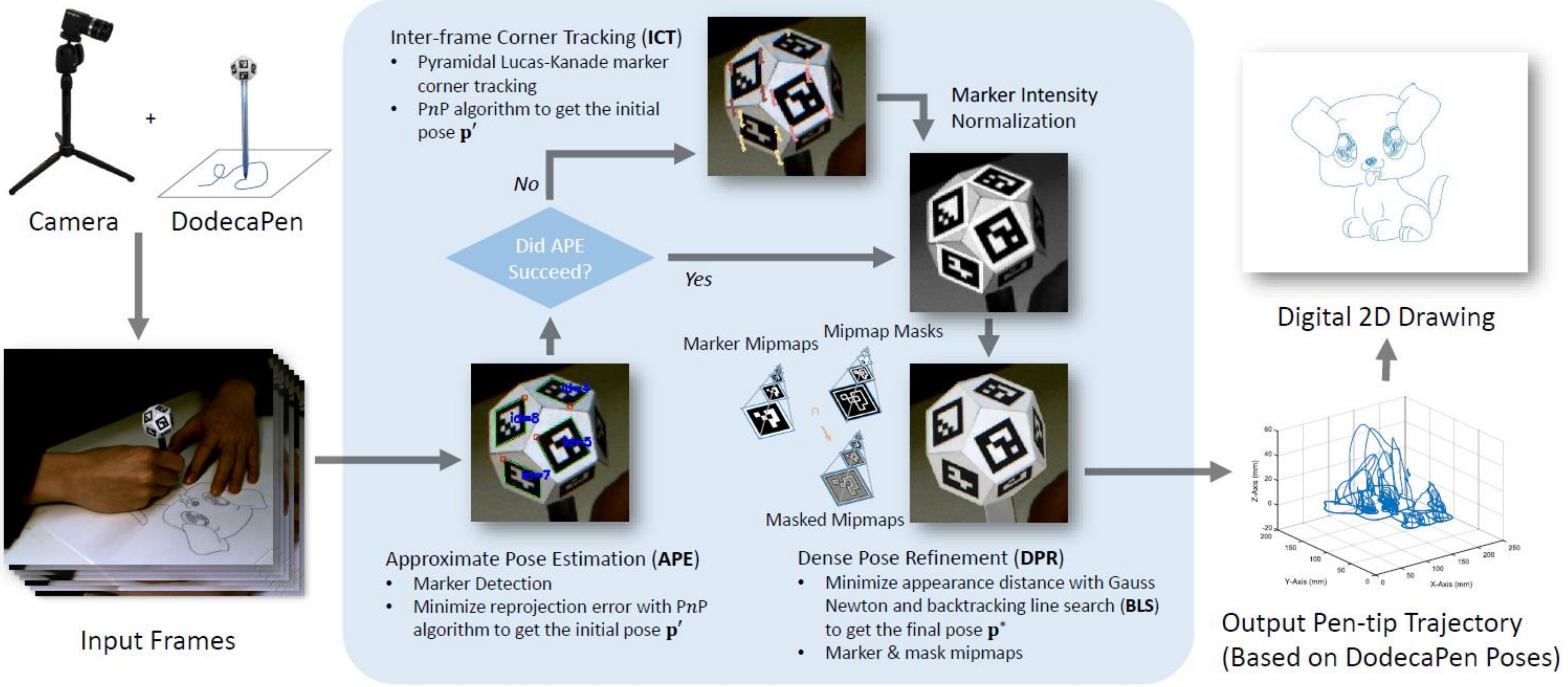
- Our system can track the 6DoF pose of a calibrated pen from a single camera with submillimeter accuracy.



(a) DodecaPen (b) Monocular video (c) DodecaPen tracking (d) Ground-truth scan

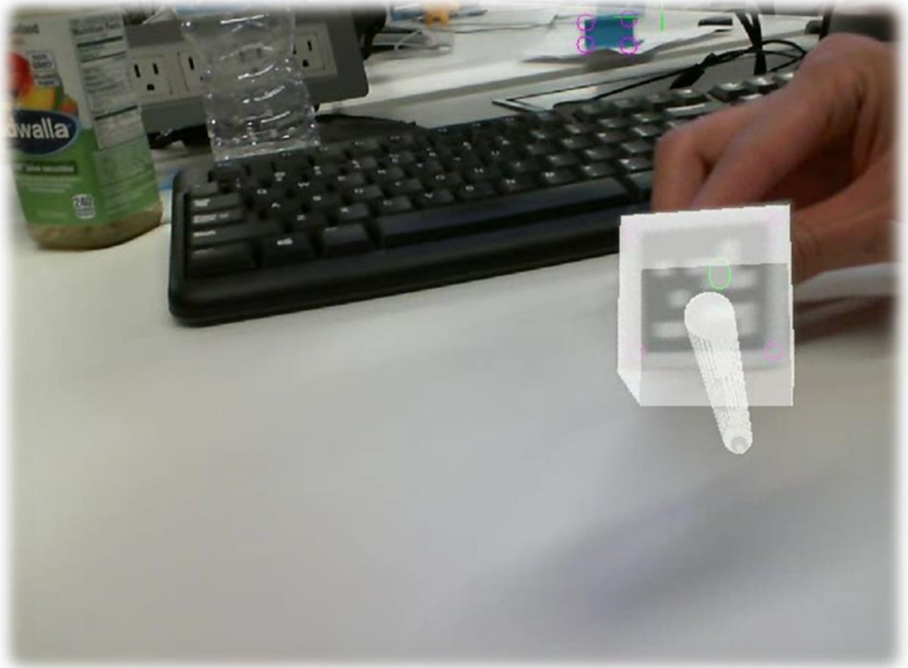
DodecaPen: Puppy

Proposed 6DoF Pose Tracking System



Why Dodecahedron?

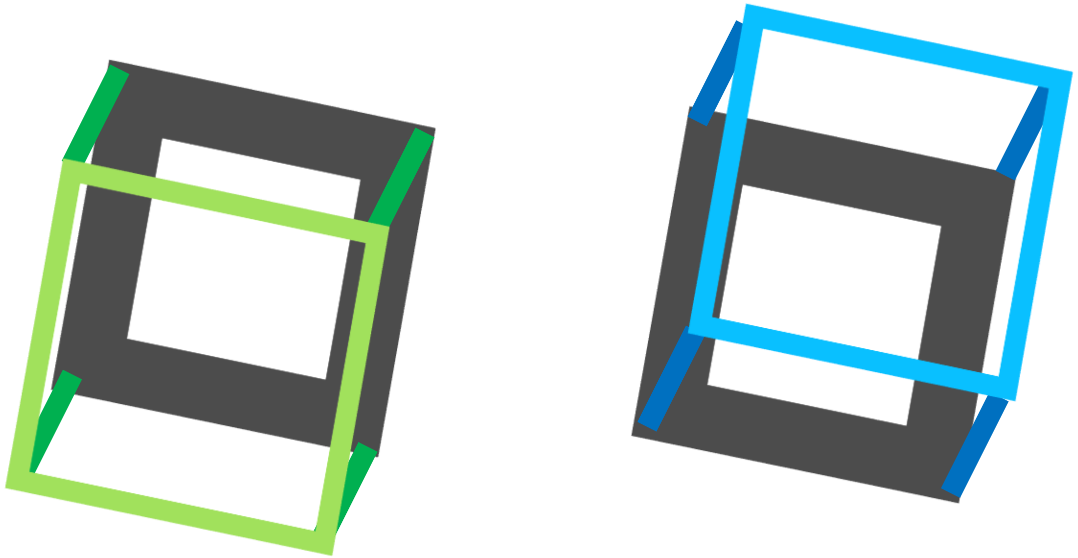
- To prevent pose jumping (due to coplanar points)



oculus

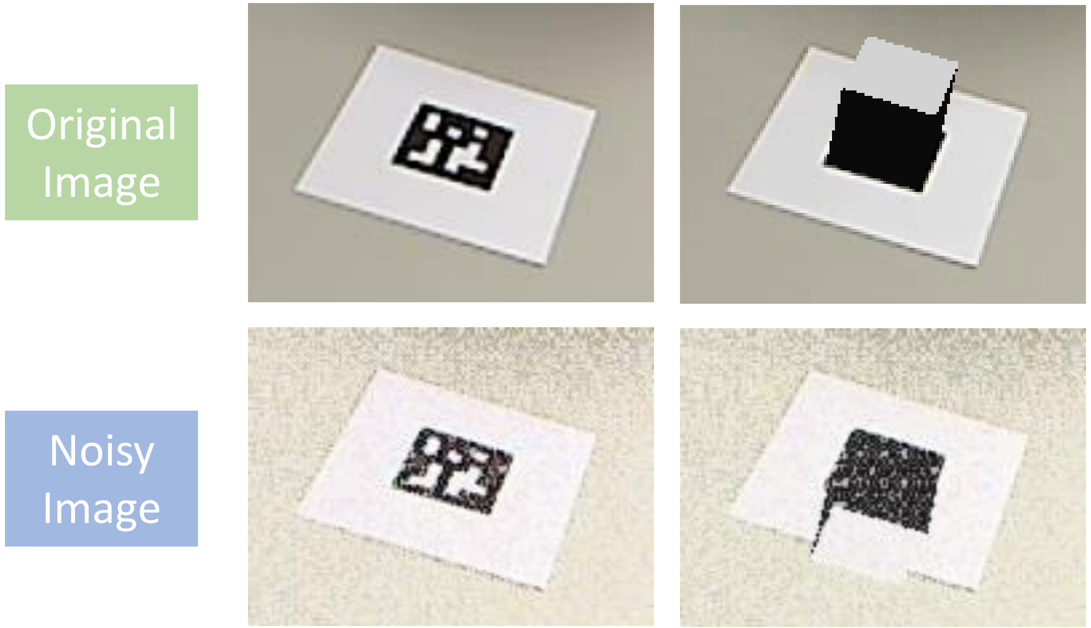
UC MERCED
UNIVERSITY OF CALIFORNIA

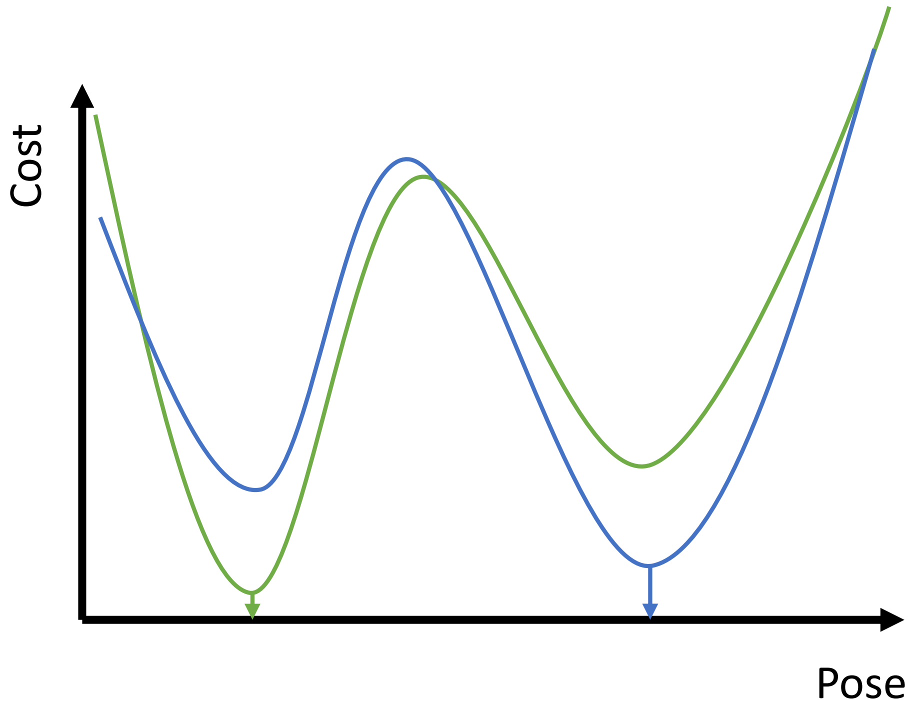
Explanation of Pose Jumping (1/2)



Explanation of Pose Jumping (2/2)

- Multiple local minima of a cost function (due to coplanar points)

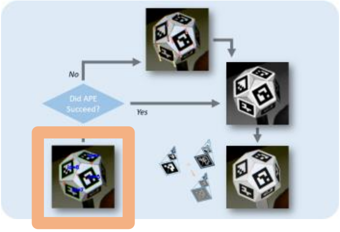




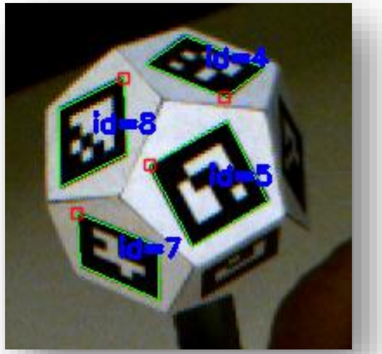
Outline

- Introduction
- Proposed System
- Experimental Results
- Conclusion

Approximate Pose Estimation (APE)



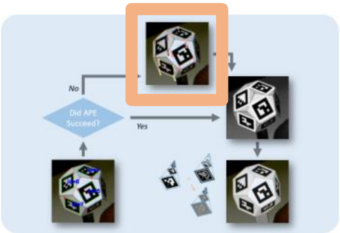
- Marker Detection
- Minimize reprojection error $E_r(\mathbf{p})$ with PnP algorithm to get the initial pose $\mathbf{p}'$

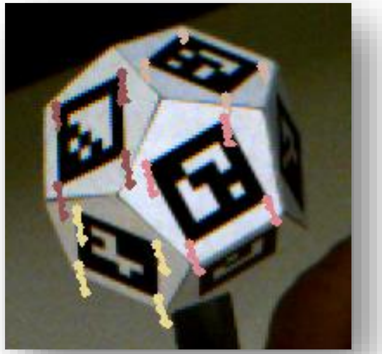


$$E_r(\mathbf{p}) = \frac{1}{n}\sum_{i=1}^{n}\left(\hat{\mathbf{u}}_i - \mathbf{u}_i(\mathbf{x}_i;\mathbf{p})\right)^2$$

$\hat{\mathbf{u}}$: detected point in the camera image
$\mathbf{x}$: point on the dodecahedron
$\mathbf{u}$: transformed $\mathbf{x}$ point in the camera image
$\mathbf{p}$: pose (including rotation matrix $\boldsymbol{R}$ and translation vector $\mathbf{t}$)

Inter-frame Corner Tracking (ICT)

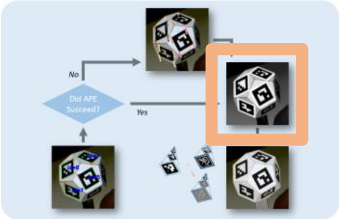




If APE does not succeed…

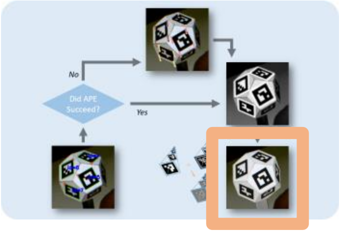
- Pyramidal Lucas-Kanade marker corner tracking
- PnP algorithm to get the initial pose $\mathbf{p}'$

Marker Intensity Normalization



- We normalize the intensity to ensure intensity invariance before minimizing the residual between the model and image.

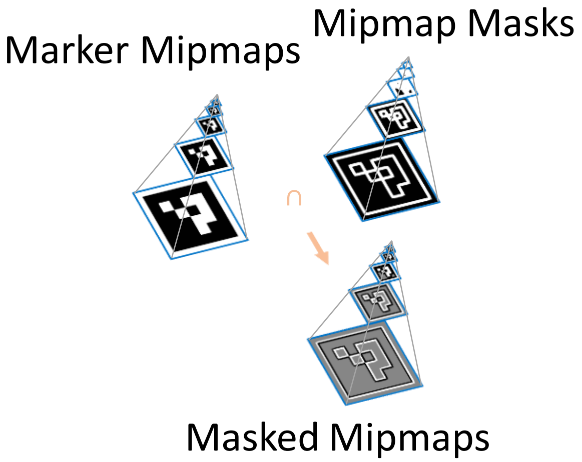
Dense Pose Refinement (DPR)



- Minimize appearance distance $E_a(\mathbf{p})$ with Gauss Newton and backtracking line search (**BLS**) to get the final pose $\mathbf{p}^*$

$$E_a(\mathbf{p}) = \frac{1}{n}\sum_{i=1}^{n}\left(I_c(\mathbf{u}_i(\mathbf{x}_i;\mathbf{p})) - O_t(\mathbf{x}_i)\right)^2$$

I_c: camera image
O_t: target object
$\mathbf{x}$: point on the dodecahedron
$\mathbf{u}$: transformed $\mathbf{x}$ point in the camera frame
$\mathbf{p}$: pose (including rotation matrix $\boldsymbol{R}$ and translation vector $\mathbf{t}$)



oculus

UCMERCED UNIVERSITY OF CALIFORNIA

Gauss-Newton Iteration

$$E_a(\mathbf{p}) = \frac{1}{n}\sum_{i=1}^{n}\left(I_c(\mathbf{u}_i(\mathbf{p})) - O_t(\mathbf{x}_i)\right)^2$$

$$\Delta\mathbf{p}^* = \underset{\Delta\mathbf{p}}{\operatorname{argmin}} \frac{1}{n}\sum_{i=1}^{n}\left(I_c(\mathbf{u}_i(\mathbf{p}' + \Delta\mathbf{p})) - O_t(\mathbf{x}_i)\right)^2$$

$$\approx \underset{\Delta\mathbf{p}}{\operatorname{argmin}} \frac{1}{n}\sum_{i=1}^{n}\left(I_c(\mathbf{u}_i(\mathbf{p}')) + \left.\frac{\partial I_c}{\partial \mathbf{p}}\right|_{\mathbf{p}=\mathbf{p}'} \Delta\mathbf{p} - O_t(\mathbf{x}_i)\right)^2$$

Vectorization $\left.\frac{\partial \mathbf{I}_c}{\partial \mathbf{p}}\right|_{\mathbf{p}=\mathbf{p}'} \equiv \mathbf{J}_c$

$E_a{}'(\mathbf{p}) = 0$

$$\mathbf{J}_c\Delta\mathbf{p} = \mathbf{O}_t - \mathbf{I}_c$$

$$\Delta\mathbf{p} = \left(\mathbf{J}_c{}^{\mathrm{T}}\mathbf{J}_c\right)^{-1}\mathbf{J}_c{}^{\mathrm{T}}(\mathbf{O}_t - \mathbf{I}_c)$$

國立臺灣大學 National Taiwan University

Backtracking Line Search (BLS)

- Gauss-Newton iteration does not always converge with a fixed step size since our least squares problem is nonlinear.

- We shrink $\Delta\mathbf{p}$ by $\Delta\mathbf{p} \leftarrow \alpha\Delta\mathbf{p}$ until it meets the *Armijo-Goldstein condition* below:

$$E_a(\mathbf{p}' + \Delta\mathbf{p}) \leq E_a(\mathbf{p}') + c\nabla E_a(\mathbf{p}')^{\mathrm{T}}\Delta\mathbf{p}$$

  - $\nabla E_a(\mathbf{p}')$is the local function gradient
  - $\alpha = 0.5\,,\, c = 10^{-4}$

Dodecahedron Calibration (DC)

- Determine the precise pose of each marker with respect to the dodecahedron
- One-time offline bundle adjustment

$$E_a(\{\mathbf{p}_j, \boxed{\mathbf{p}_k}\}) = \sum_i \sum_j \sum_k \left(I_c\left(\mathbf{u}_i(\mathbf{x}_i; \mathbf{p}_j; \mathbf{p}_k)\right) - O_t(\mathbf{x}_i)\right)^2$$

I_c: camera image
O_t: target object
$\mathbf{x}$: point on the dodecahedron
$\mathbf{u}$: transformed $\mathbf{x}$ point in the camera frame
$\mathbf{p}$: dodecahedron pose
$\mathbf{p}$: marker pose

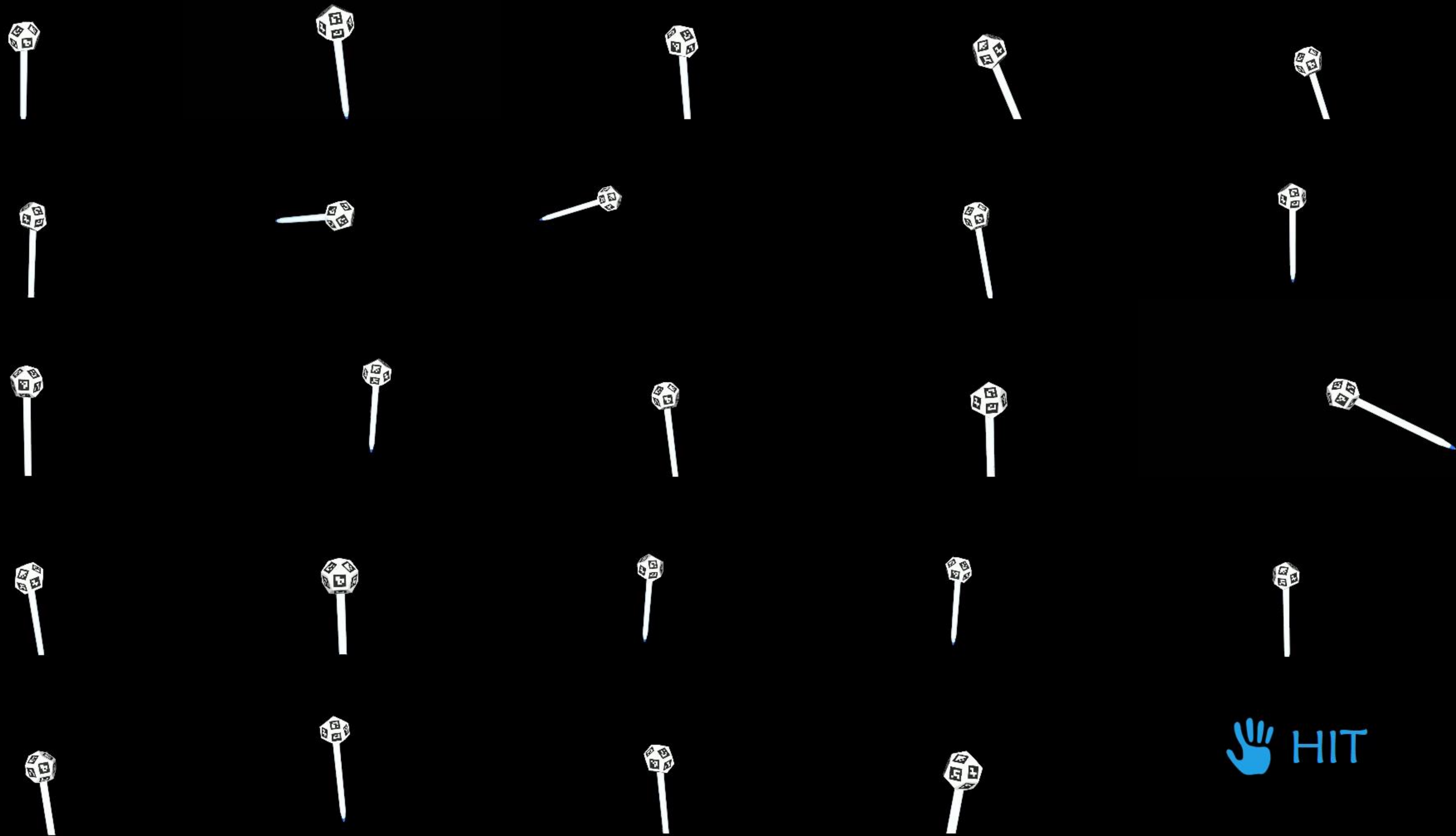
Outline

- Introduction
- Proposed System
- **Experimental Results**
- Conclusion

Evaluation with Synthetic Data

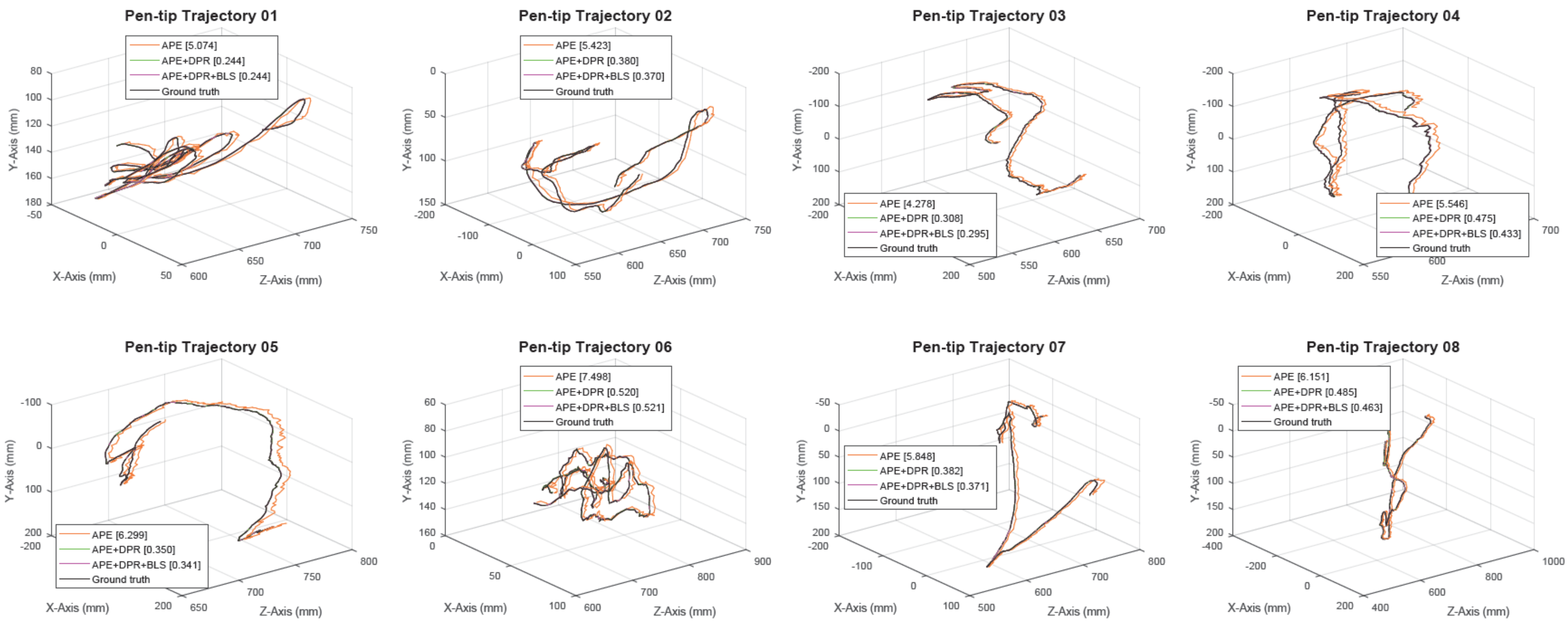
- We generate synthetic image sequences with 24 motion patterns of the virtual DodecaPen
  - Each sequence consists of 301 frames with the same resolution (1280*1024) and intrinsics as our real camera.

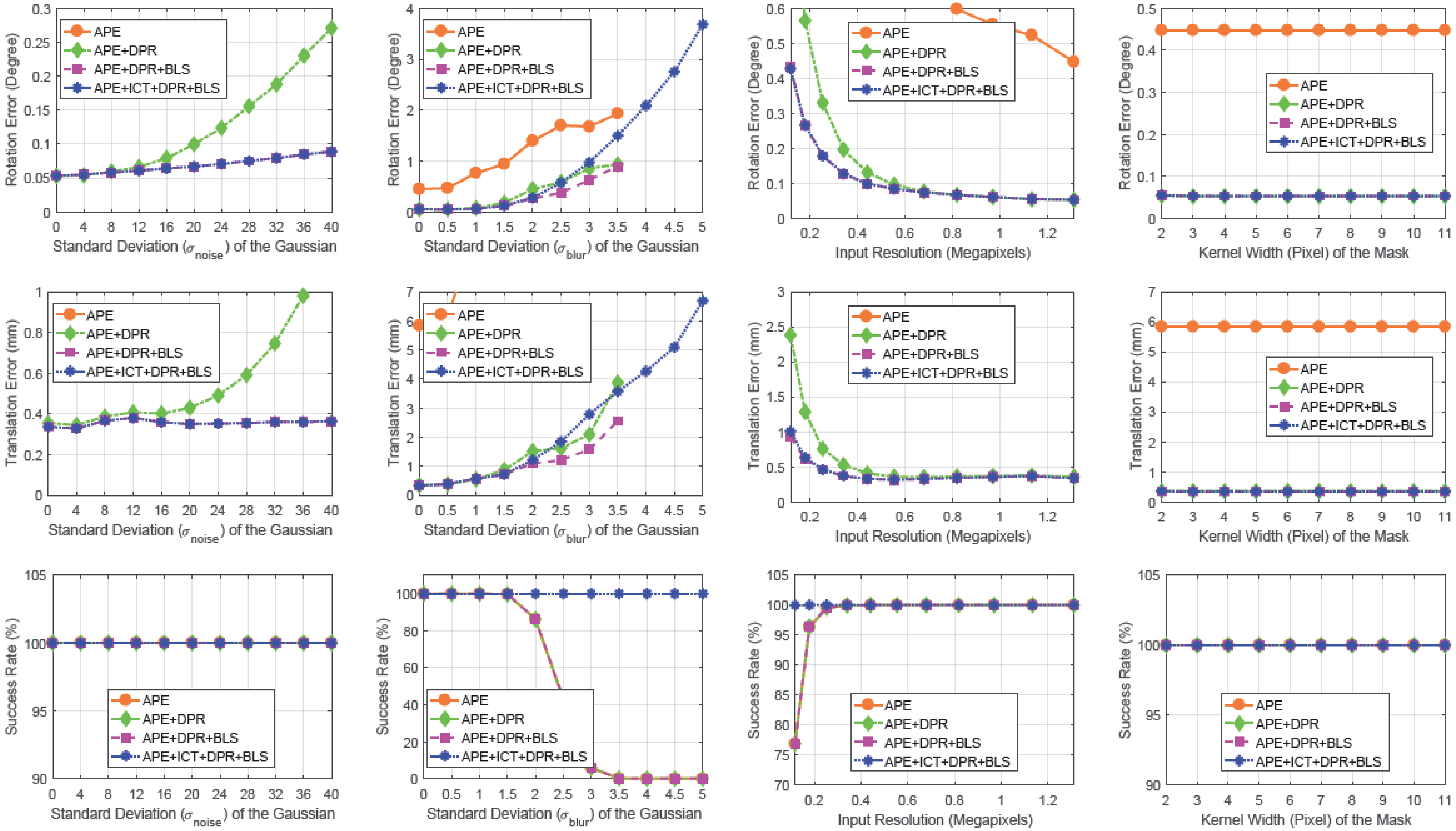
- We further evaluate the proposed approaches under varying shot noise, spatial blur, camera resolutions, and mask kernel widths.



HIT

Pen-tip Trajectories

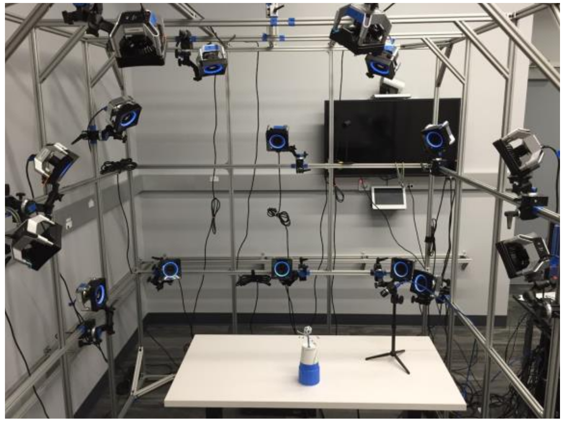
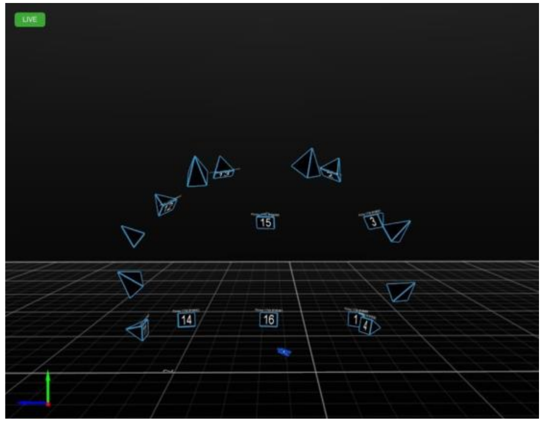
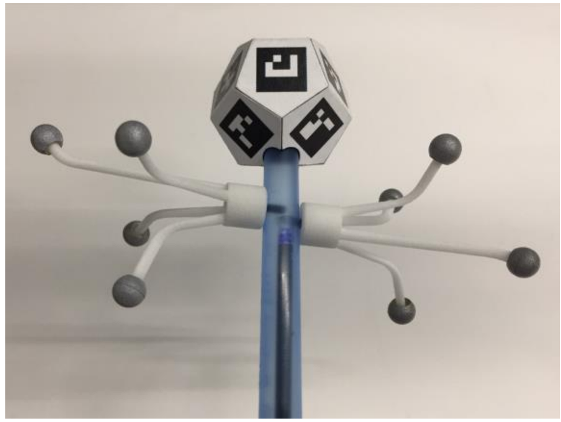
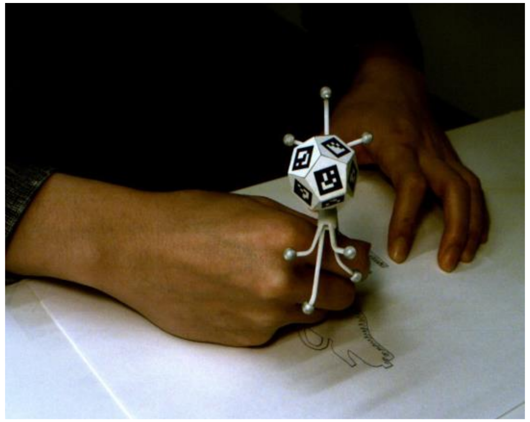


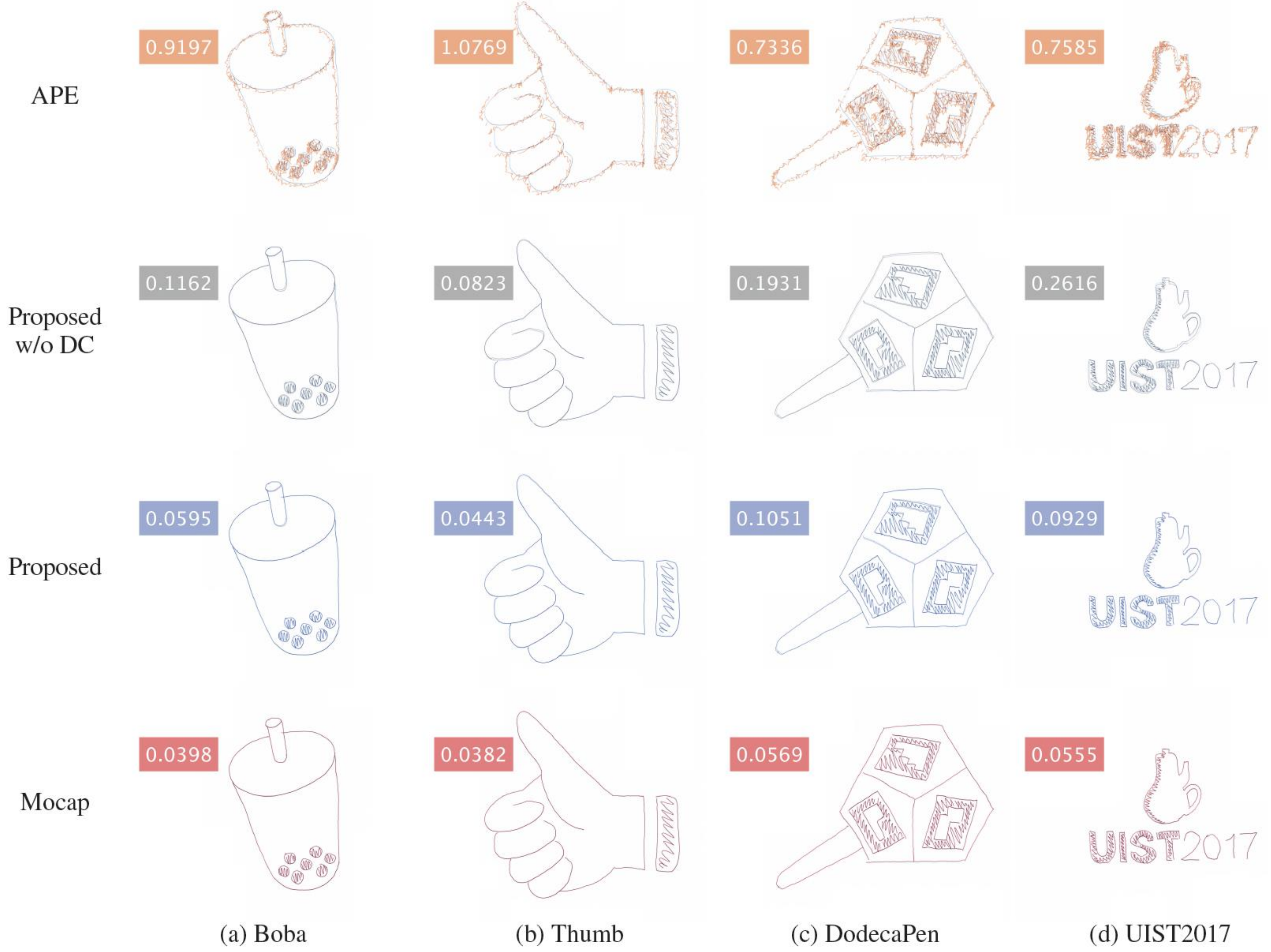


(a) Shot noise (b) Spatial blur (c) Camera resolution (d) Mask kernel width

Evaluation with Real Data

- Four real drawings
- Compare with a motion capture system
  - Constructed with 16 OptiTrack Prime 17W (1.7 megapixels, 70 degrees field-of-view) cameras

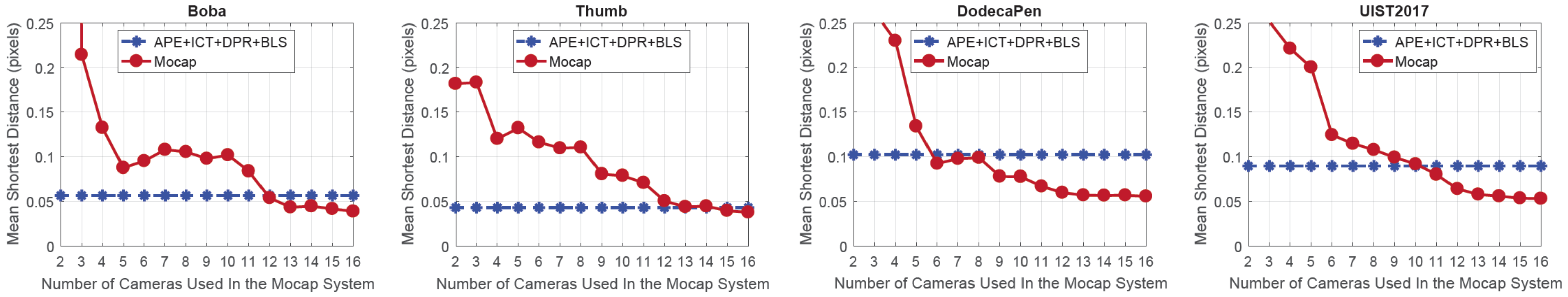







APE
Proposed w/o DC
Proposed
Mocap
0.9197
1.0769
0.7336
0.7585
0.1162
0.0823
0.1931
0.2616
0.0595
0.0443
0.1051
0.0929
0.0398
0.0382
0.0569
0.0555
UIST2017
(a) Boba
(b) Thumb
(c) DodecaPen
(d) UIST2017

DodecaPen VS. Mocap

- The accuracy of the proposed method is comparable to a motion capture system with 10 active cameras.
- TW$ 10,000 VS. TW$ 1,000,000



Outline

- Introduction
- Proposed System
- Experimental Results
- Conclusion

Conclusion

- 6DoF tracking using a set of readily available and easy-to-assemble components
- Sub-millimeter accurate
- Single camera pose estimation can be fast enough and robust enough for drawing in 2D, 3D and in VR

Potential AI Research Topics

- General Object Pose Tracking