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

Accurate 6DoF Object Pose Estimation and Tracking

A Dissertation Defense

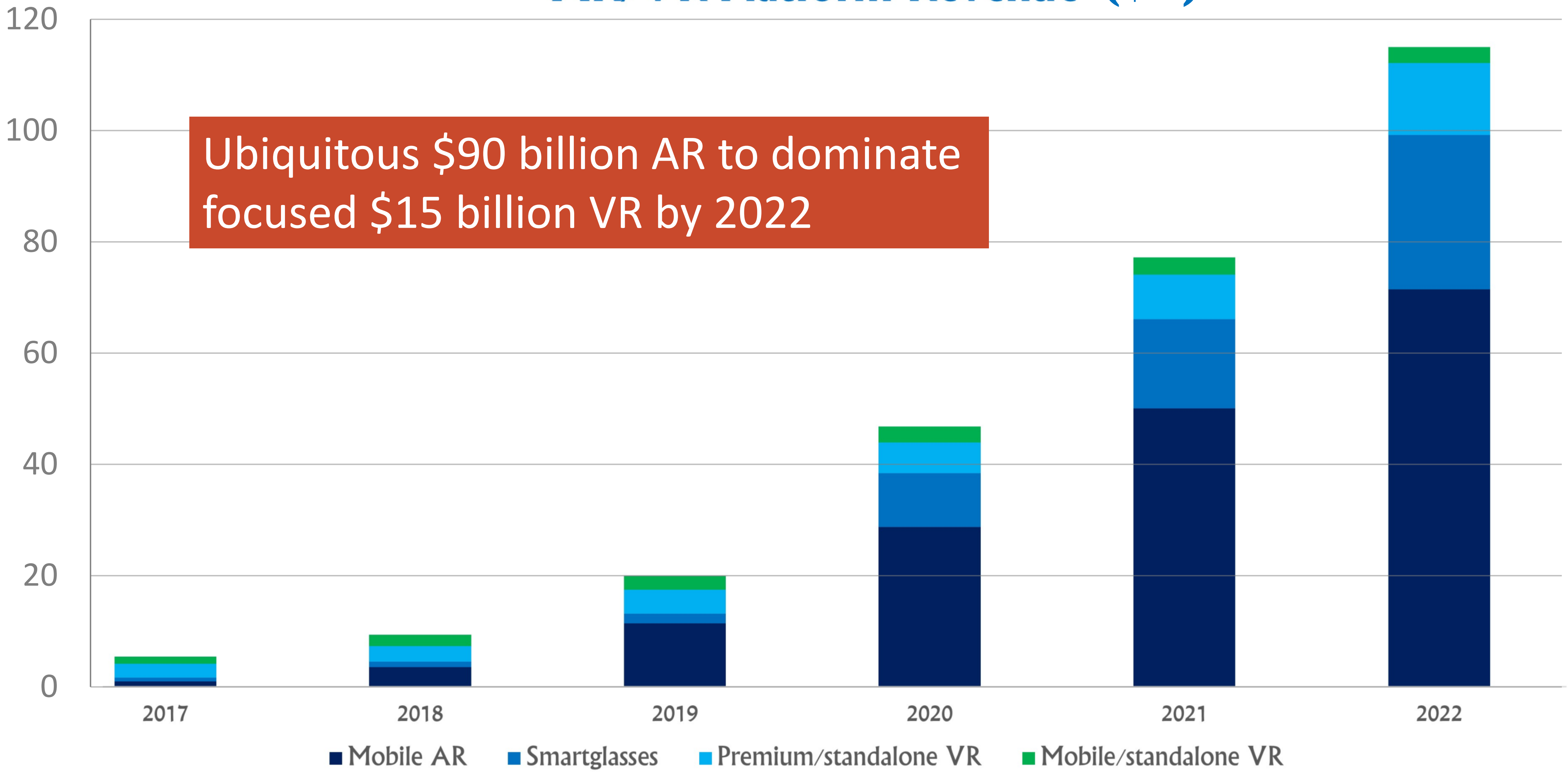
by

Po-Chen Wu

Advisor: Dr. Shao-Yi Chien

Co-Advisor: Dr. Ming-Hsuan Yang

AR/VR Platform Revenue ($B)



Digi-Capital™

Augmented Reality

https://giphy.com/gifs/adweek-place-ar-4R63eQx8wyEda

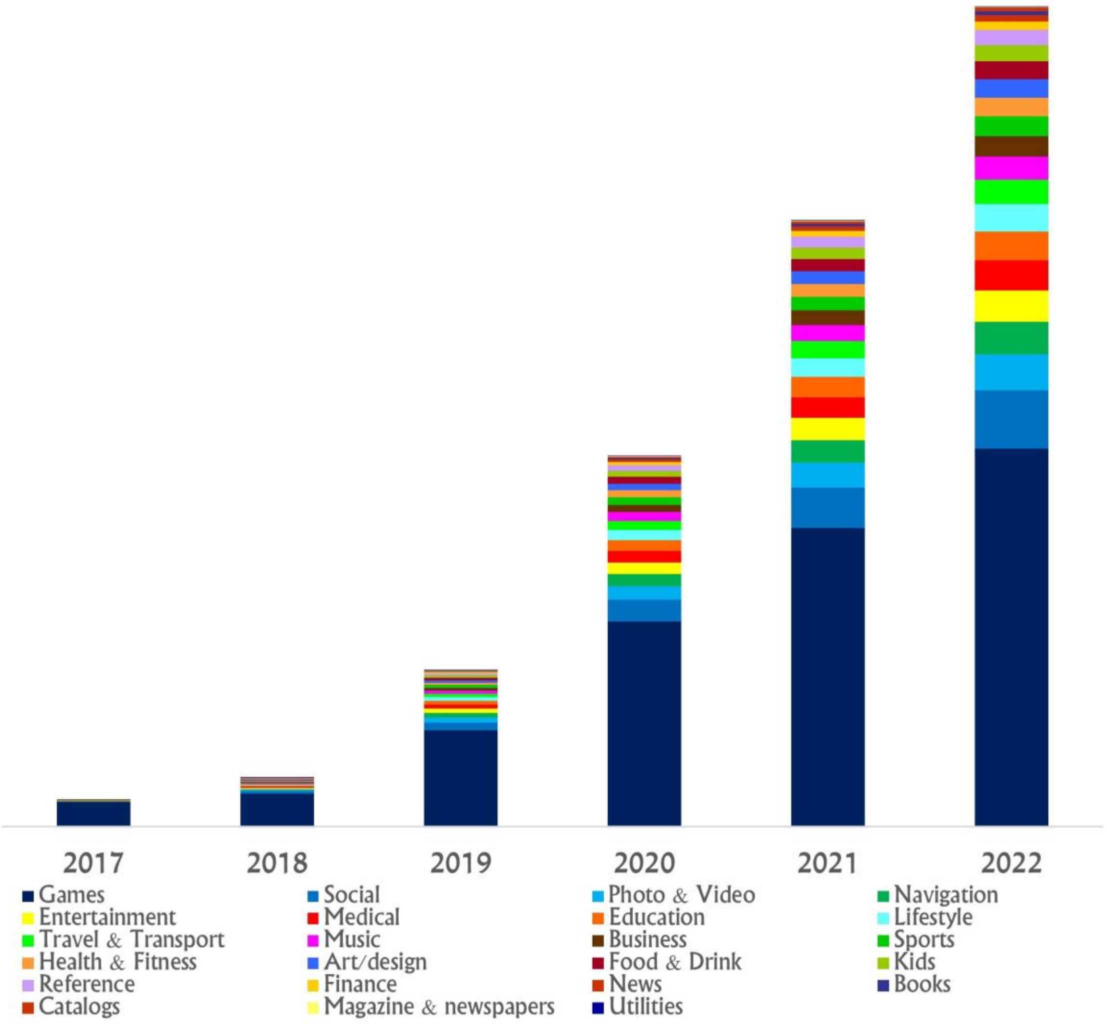
Virtual Reality

https://cdn.makeuseof.com/wp-content/uploads/2017/01/oculus-touch-gestures-gif.gif

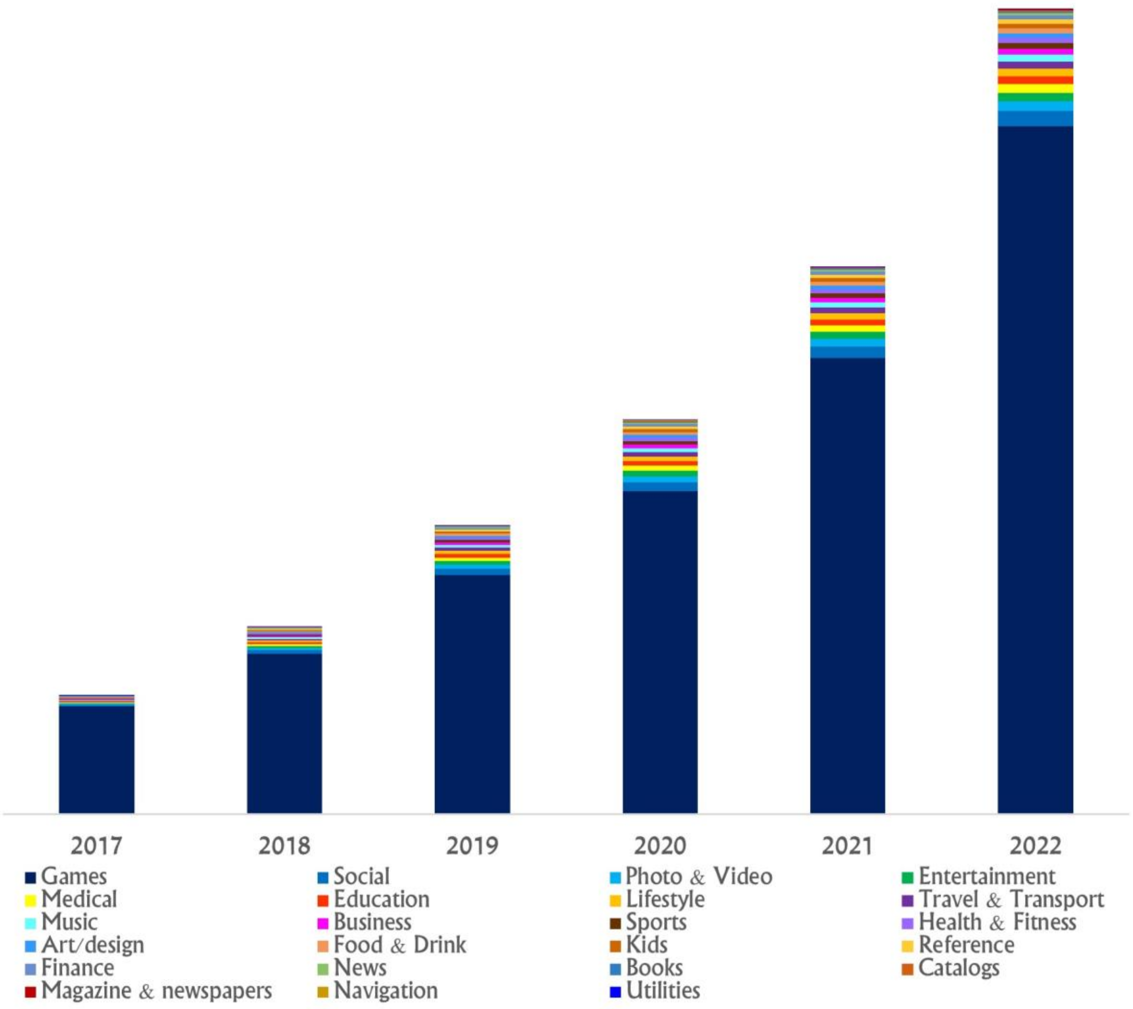
AR/VR App Store Category Revenue (IAP/Premium)

(Note: scales on both charts are not the same)

AR Category (IAP/Premium) Revenue ($90B)



VR Category (IAP/Premium) Revenue ($15B)



Digi-Capital™

Google

https://inhabitat.com/ecouterre/your-surgeon-could-be-using-google-glass-in-the-operating-room/

https://www.theverge.com/2016/11/1/13480840/google-tango-lenovo-phab-2-pro-apps-games-release-date

Facebook

https://www.independent.co.uk/life-style/gadgets-and-tech/news/oculus-rift-price-headset-and-computers-that-can-run-it-begin-at-1499-a6863676.html

http://immediatefuture.co.uk/blog/facebook-launches-ar-studio/

Apple

https://www.tomsguide.com/us/apple-ar-glasses-tim-cook,news-25964.html

https://mashable.com/2017/06/05/apple-arkit-hands-on/

Microsoft

https://www.geeky-gadgets.com/microsoft-windows-10-vr-headset-22-11-2016/

http://fortune.com/2017/02/21/microsoft-hololens-update-delay/

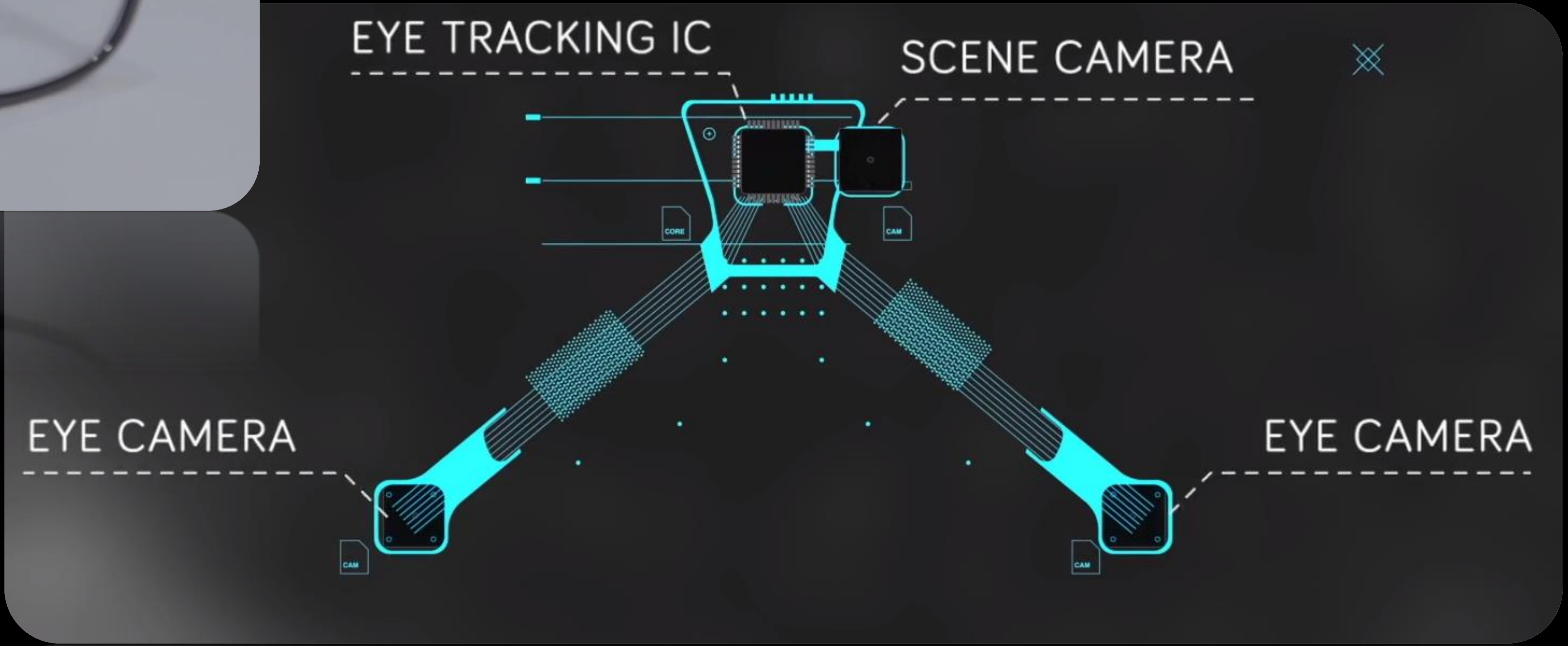
Magic Leap

https://www.engadget.com/2018/02/13/magic-leaps-ar-headsets-price-nba-deal/

https://www.engadget.com/2017/12/20/magic-leap-one-details-questions-dont-know/

GANZIN

https://www.youtube.com/watch?v=d9_kYHbEI5w



https://www.youtube.com/watch?v=d9_kYHbEI5w

(02) 3366-3668

Digi-Capital™ AR/VR Leaders*

** Includes funded/exited startups and selected corporates*

Advertising/marketing
adle, AdsReality, ADVR, Submersive, Fake Love, Omnivirt, Immersal, Real Vision, ZeniAd, retinad, AdVir, blippar, Plattar, Augment, Personas, advrty, immersv, vertebrae, VIRE, Xperiel, zappar

Art/design
Mindesk, iris, artomatix

Business
cluster, VIRTUALITICS

Distribution
ConstructVR, Wevr, WEARVR, Sketchfab, UNIMERSIV, Rock VR

eCommerce
omagine, CAPASITY, VIVIDWORKS, InContext Solutions, spaceview, hexa, tylko

Education
ARA, nearpod, cerevrum, serious labs, mativision, Immersive Education, sdk, LVR, MEL Science, CURISCOPE, JigSpace, Learningbox, VISCOPIC, OpenSesame, LABSTER, STRIVR, VPS, DISCOVR LABS, BOULEVARD, zSpace

Enterprise/B2B
UPSKILL, Mure, Hyper Fair, SCOPE AR, taqtile, IMMERSION, soluis, FIELDBIT, InsiteVR, SPACESYS

Entertainment
Madefire, TRIPP, SKYLIGHTS

Games
Rec Room, Against Gravity, Anima VR, Applied Ideas, Boom, bublar, Directive Games, TEQ, ESCALATION, Fast Travel Games, Freeform Labs, Munzee, Multiverse, Midas Touch Interactive, Resolution, fuzzycube, Harmonix, metaverse, KENZAN, Kingdom of Castles, CCP, Cyberpong VR, sólfar, Survios, Sliver.tv, Survive, Superstar, Osmo, Polyarc, Realiteer, Niantic, nDreams, Owlchemy Labs, playful, PLAY Fusion, Shortfuse, Temple Gates, umbra, Starbreeze, VRANI, PROXY42, Kite & Lightning, ROBLOX, VITO, iNK'LE, WARDUCKS

Kids
QuiverVision

Lifestyle
Around Media, Habiteo, GeoCV, Vieweet, iStaging, pixie, roOomy, urbanbase, MODSY

Location based
igloo, Enter Space, Havson, VRSTUDIOS, IMAX VR, NOMADIC, Dreamscape, VRcade, hyperverse, LeKe VR, ZERO Latency, TWO BIT CIRCUS, Spaces, Force Field

Medical
aira, appliedVR, AUGMEDIX, BIOLUCID, Luminate, echopixel, Medical Realities, Surgical Theater, Kavision, psious, Fearless, mimesys, marion surgical, orosim, OVID, mindmaze, oriense

Music
TheWaveVR

Navigation
Aero Glass, NAVISENS, mapbox, PARACOSM, WAYRAY

News
87870.com, HAMMER & TUSK, UPLOAD

Peripherals
Remoria, TEGway, Xenxe, Ximmerse, 9ONE, RAZER, VRZoom, VAQSO, UMBO, OSSIC, Myo, LIGHTFORM, SUBPAC, POLARIANT, Tactical Haptics, YOST, ultrahaptics, aeroVR, nod, here, CyberGlove, LEAP, HYVE, Omnifinity, haptics, Cyberith, KAT VR, Neurodigital

Photo/video
FOV, Yusion, scandy, 360fly, ALCHEMY VR, Archiact, Blend, briskeye, bubl, Bitmovin, Construct Studio, dRi, EEVO, Felix & Paul, LYTRO, entrypoint, GoPro, GIROPTIC, Infinite Mixed Reality, InstaVR, eko, JAUNT, Kite & Lightning, kogeto, Kolor, livelike, LUCID, Lumiere VR, matterport, LUNA, METTLE, nctech, NEXTVR, NGCODEC, OZO, otoy, Panocam, PARABLE, PENROSE, REALITIES.IO, RYOT, Spacevr, SpinCle, Sureal, The Future, Tiplt, ORAH, VISBIT, VISUALPATHY, VOKE, Playhouse, zLense, EMERGENT, Littlstar

Productivity

Smartglasses
Google, ATHEER, PEPPER, DAQRI, EPSON, Cdi, WaveOptics, KOPIN, magic leap, AVEGANT, CMOAR, Microsoft, Real View, Lenovo, Immy, Vuzix, ODG, SULON, Meta, SONY, kikaku, mira, intel, recon

Solutions/services
ARHASS, Digital Domain, LOOM.AI, pixelbug, spotscale, Nomo, DoubleMe, DVERSE, FISHBOWL VR, HOLOGROUP, DIAKRIT, 8i, 8WALL, BinaryVR, ARHT Media, catchoom, BYOND, cognitive3D, Dacuda, DataMesh, Apmetrix, GAUDIO, iris, MELODY VR, OVA, snobal, smartvizx, VECTARY, mindshow, wrnch, SIMPLYGON, Mobeck

Social
FlirtAR, PLUTO, NEOS, SPACEOUT VR, CLEVR, Decentraland, Surreal, High Fidelity, VRCHAT, TIMEFIRE VR

Sports
HOLODIA, FIRSTVISION, eon sports, SoccerDream, BEYOND SPORTS

Tech
VicoVR, 3DLOOK, actronika, AMD, Apple, amazon, AirPods, Aireal, crunchfish, CRYTEK, CUBICMOTION, DIGILENS, Dirac, DOLBY, DYSONICS, EONITE, Epic Games, GaitUp, gestigon, Google, Infinity, Lumus, Manomotion, MOGEES, Movidius, neurable, nGRAIN, nitero, NEURXEN, NVIDIA, occipital, ptc, Reach, SMI, SoftKinetic, Varjo, THE EYE TRIBE, Third Eye, unity, UNIVRSES, worldviz, VisiSonics, xPerception, SENNHEISER, IMPROBABLE, insightness, antilatency, THRIVE AUDIO, BrokenColors

Travel/transport
ZeroLight, STURFEE, timelooper, GAMAR, CAR360

Utilities
BAGEL LABS, bigscreen, Seene, visual camp, JANUSVR, nurulize, Vizor, wakingapp, HASHPLAY

VR headset
NORAH, alcatel, altergaze, ARCHOS, ZEISS, 暴风魔镜, VUZIX, mi, facebook, FOVE, Google, XINGEAR, LG, airVR, MERGE, Microsoft, Give Vision, POWIS, RAZER, SONY, htc, Immersion Vrelia, STARBREEZE, Dee Poon, wearality, evomade, VALVE, SAMSUNG, sensics, EYEDAK

Pose Estimation + Rendering

https://giphy.com/gifs/3d-LGP31CCUCylbO

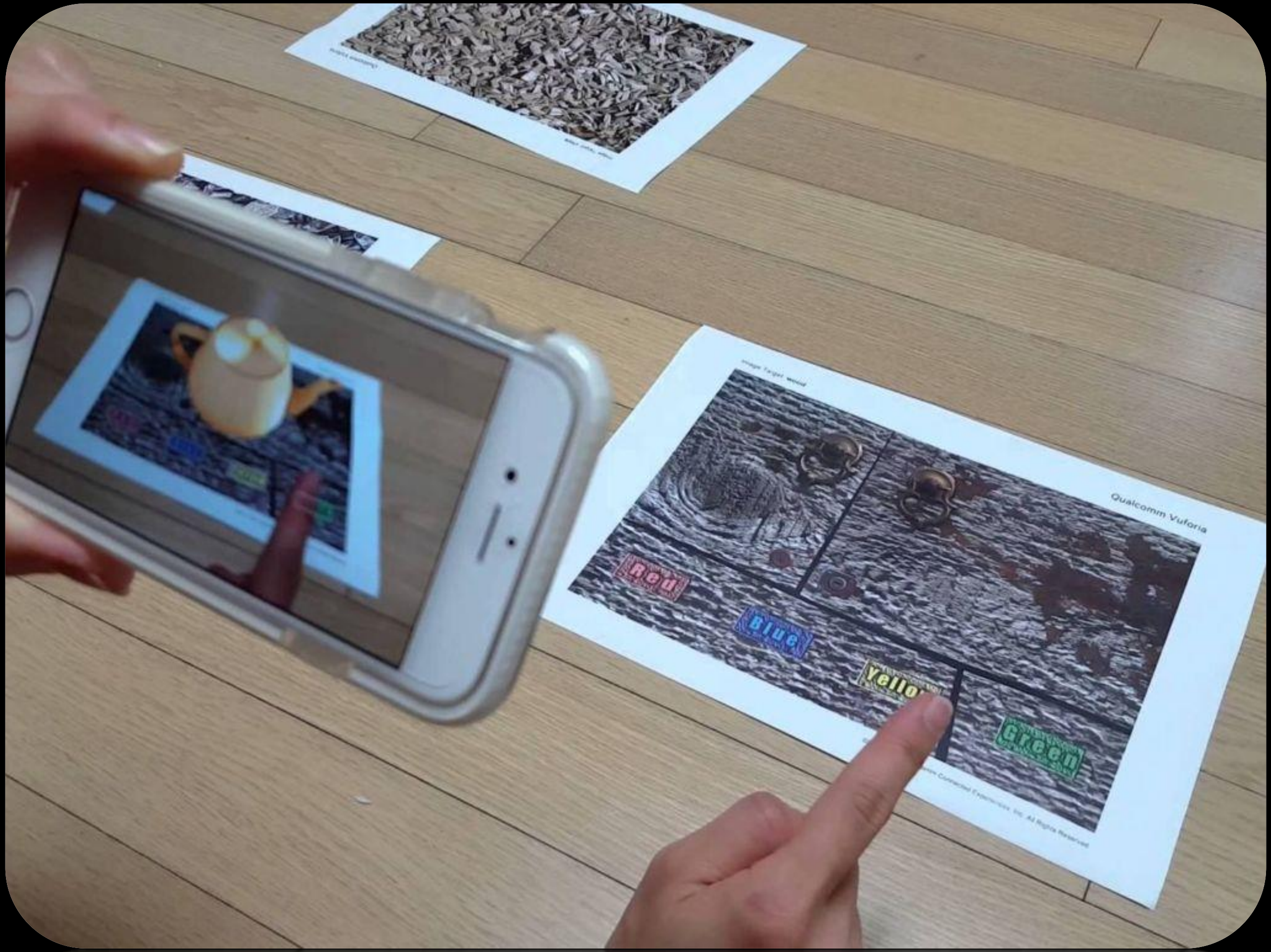
Six Degrees of Freedom (6DoF) (Rigid Body) Object Pose

Position (3DoF)

+

Orientation (3DoF)

Problem Definition

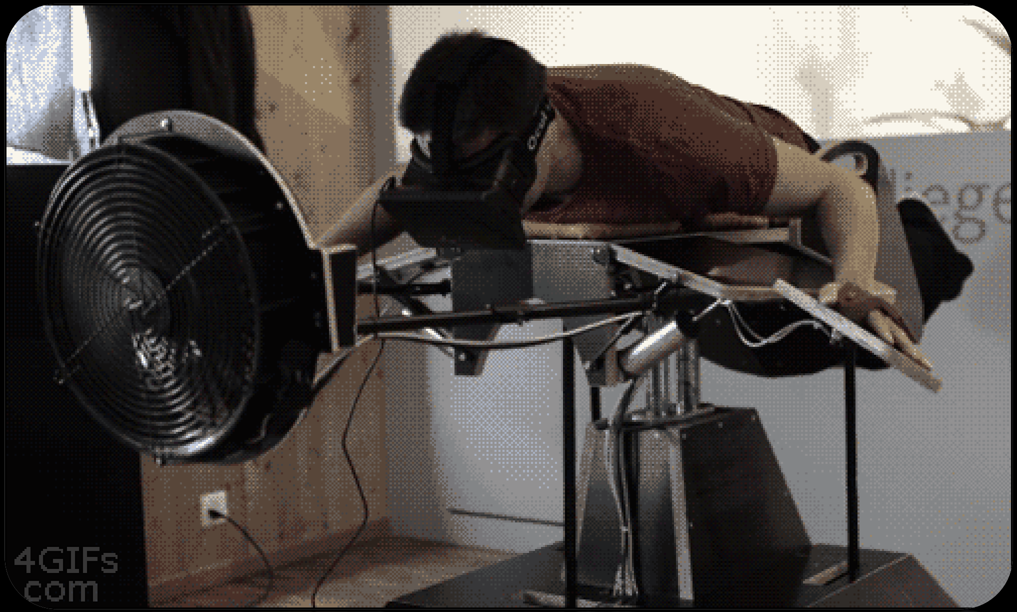


https://www.zuehlke.com/blog/apple-arkit-augmented-reality-erhaelt-schub/

AR

https://www.microsoft.com/en-us/hololens

VR



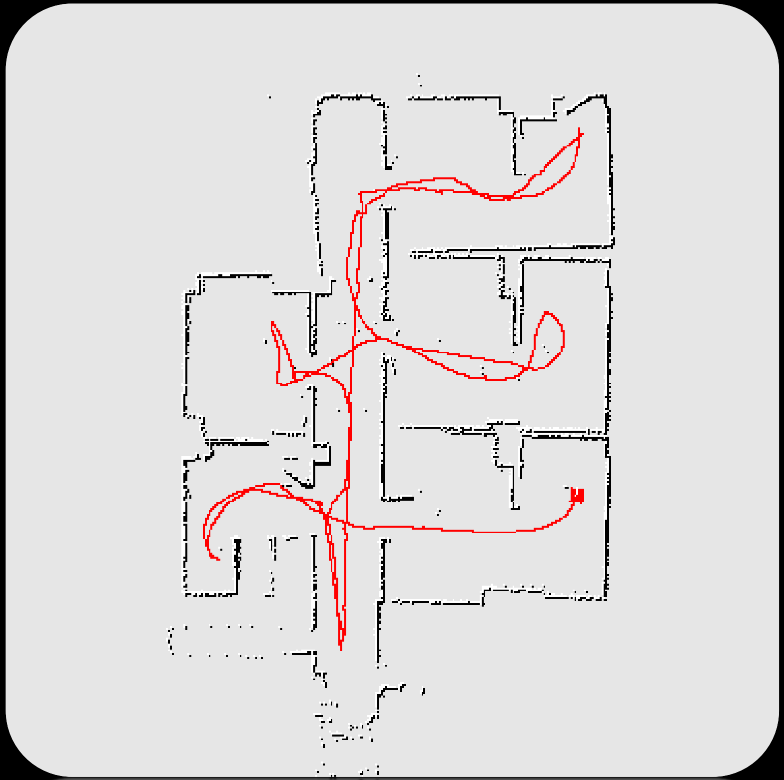
https://giphy.com/gifs/3d-simulator-bird-bl09MDvT4JIWc

6DoF Object Pose

- Model-Based Pose Estimation
- Model-Based Pose Tracking

6DoF Camera Pose

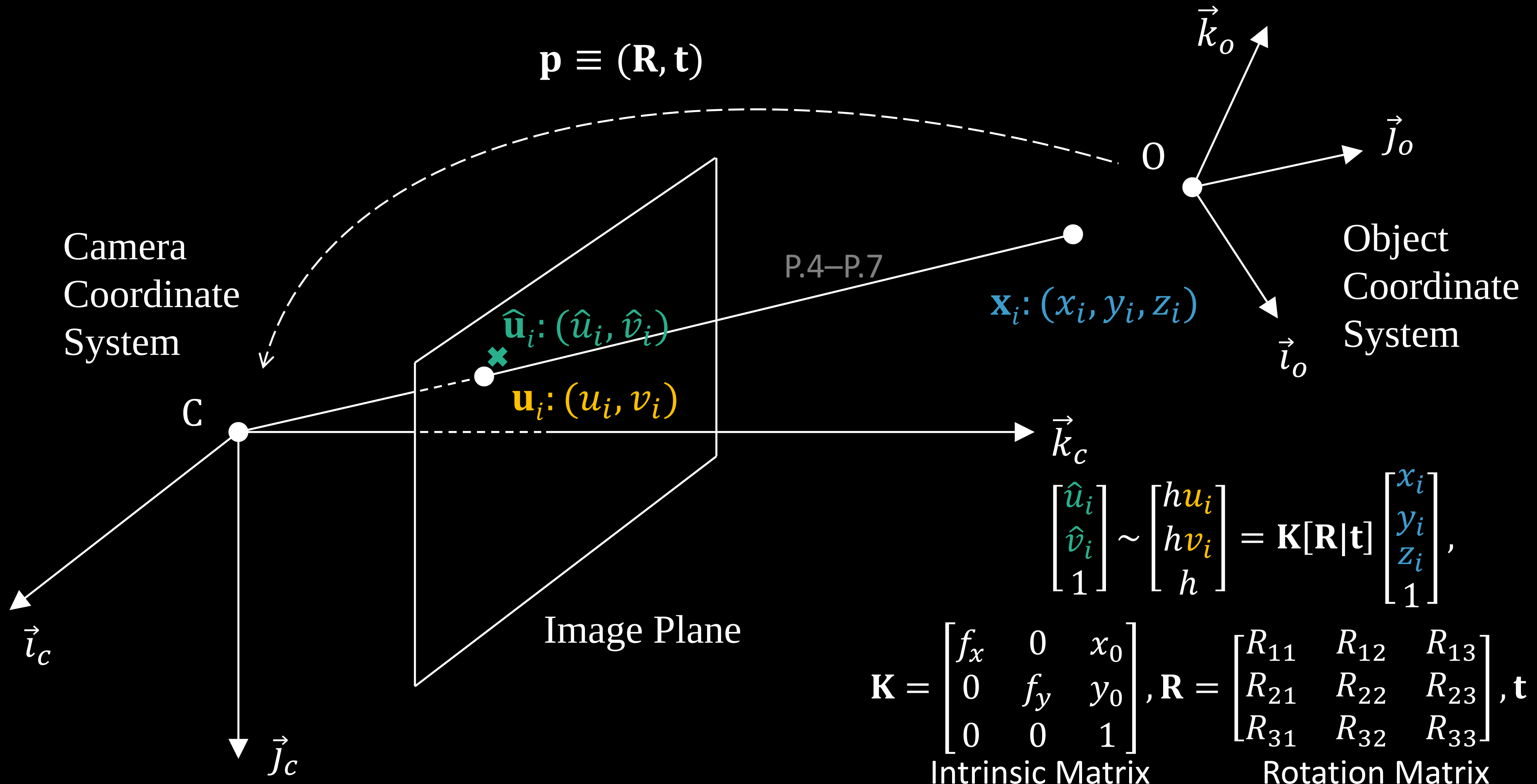
- Image-Based Camera Localization
- Simultaneous Localization and Mapping (SLAM)



http://www.societyofrobots.com/robotforum/index.php?topic=16714.0

P.13–P.14

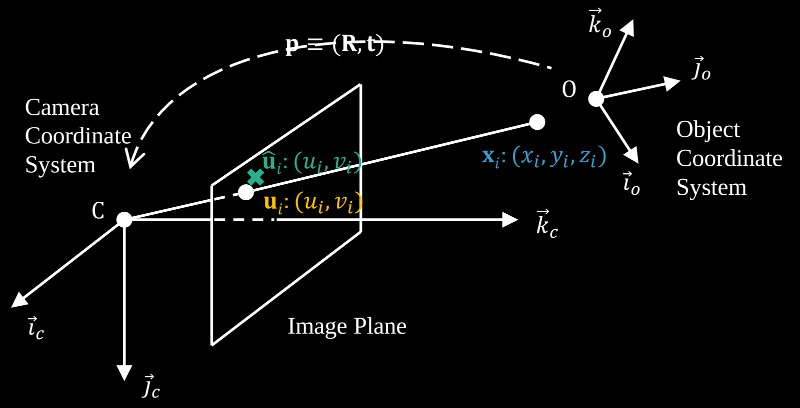
Camera Projection Model



$$\begin{bmatrix} \hat{u}_i \\ \hat{v}_i \\ 1 \end{bmatrix} \sim \begin{bmatrix} hu_i \\ hv_i \\ h \end{bmatrix} = \mathbf{K}[\mathbf{R}|\mathbf{t}] \begin{bmatrix} x_i \\ y_i \\ z_i \\ 1 \end{bmatrix},$$

$$\mathbf{K} = \underbrace{\begin{bmatrix} f_x & 0 & x_0 \\ 0 & f_y & y_0 \\ 0 & 0 & 1 \end{bmatrix}}_{\text{Intrinsic Matrix}}, \mathbf{R} = \underbrace{\begin{bmatrix} R_{11} & R_{12} & R_{13} \\ R_{21} & R_{22} & R_{23} \\ R_{31} & R_{32} & R_{33} \end{bmatrix}}_{\text{Rotation Matrix}}, \mathbf{t} = \underbrace{\begin{bmatrix} t_x \\ t_y \\ t_z \end{bmatrix}}_{\text{Translation Vector}}$$

Objective Functions



- Reprojection error

$$E_r(\mathbf{p}) = \frac{1}{n}\sum_{i=1}^{n}\left((\hat{u}_i - u_i)^2 + (\hat{v}_i - v_i)^2\right)$$

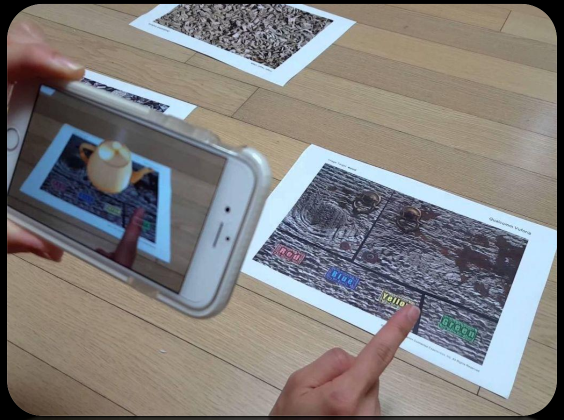
$\hat{\mathbf{u}}_i: (\hat{u}_i, \hat{v}_i)$ is the observed point

- Appearance distance

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

or

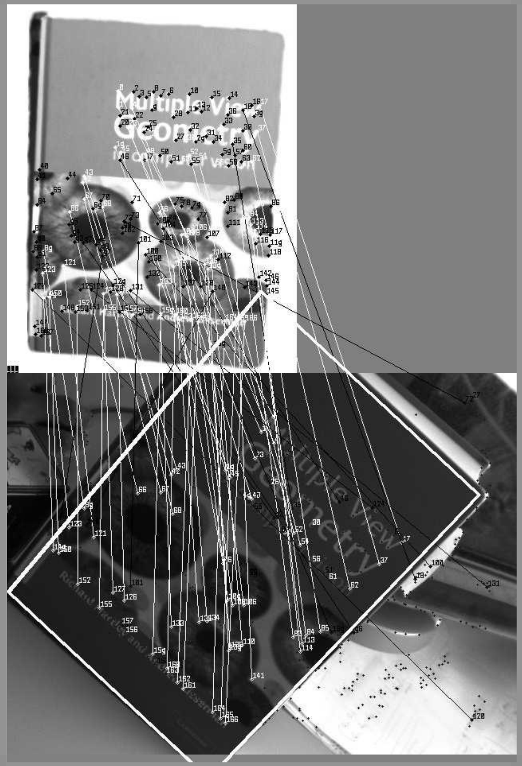
$$E_a(\mathbf{p}) = \frac{1}{n}\sum_{i=1}^{n}\left(I_c(\mathbf{u}_i) - O_t(\mathbf{x}_i)\right)^2$$



https://www.zuehlke.com/blog/apple-arkit-augmented-reality-erhaelt-schub/

Feature-Based Approaches [1,2,3]

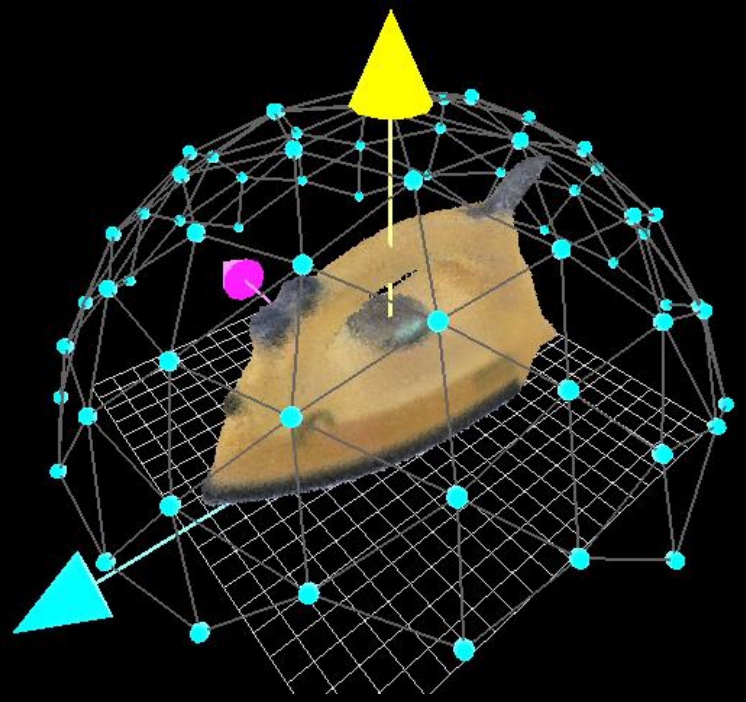
- Feature detection and matching
  - SIFT, SUFT, FAST, ORB, …
- Outlier removal
  - RANSAC, PROSAC, …
- Perspective-n-point (PnP) Algorithm
  - EPnP, OPnP, UPnP, …
  - $E_r(\mathbf{p}) = \frac{1}{n}\sum_{i=1}^{n}||\hat{\mathbf{u}}_i - \mathbf{u}_i||^2$



[1] Lepetit et al., "Point matching as a classification problem for fast and robust object pose estimation," CVPR, 2004.
[2] Collet et al., "Object Recognition and Full Pose Registration from a Single Image for Robotic Manipulation," ICRA, 2009.
[3] Tang et al., "A Textured Object Recognition Pipeline for Color and Depth Image Data," ICRA, 2012.

Direct Approaches [1,2]

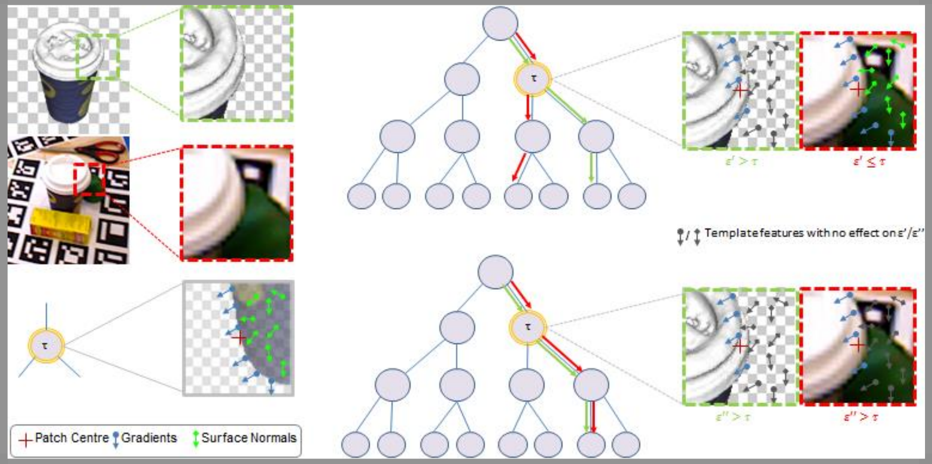
- Finding the best fit from numerous pre-determined candidates
  - $E_a(\mathbf{p}) = \frac{1}{n}\sum_{i=1}^{n}|I_c(\mathbf{u}_i) - O_t(\mathbf{x}_i)|$
  - $E_a(\mathbf{p}) = \frac{1}{n}\sum_{i=1}^{n}\left(I_c(\mathbf{u}_i) - O_t(\mathbf{x}_i)\right)^2$



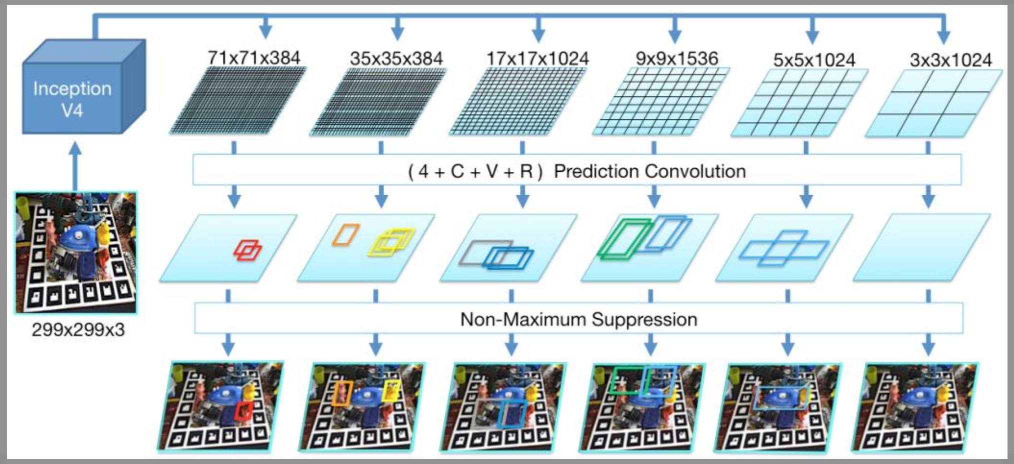
[1] Hinterstoisser et al., "Model Based Training, Detection and Pose Estimation of Texture-Less 3D Objects in Heavily Cluttered Scenes," ACCV, 2012.
[2] Tseng et al., "Direct 3D Pose Estimation of a Planar Target," WACV, 2016.

Learning-Based Approaches

- Random forest [1]



- Deep neural network [2,3]



[1] Tejani et al., "Latent-Class Hough Forests for 3D Object Detection and Pose Estimation," ECCV, 2014.
[2] Kehl et al., "SSD-6D: Making RGB-Based 3D Detection and 6D Pose Estimation Great Again," CVPR, 2017.
[3] Rad et al., "BB8: A Scalable, Accurate, Robust to Partial Occlusion Method for Predicting the 3D Poses of Challenging Objects without Using Depth," ICCV, 2017.

Objective

Accurate

Real-time

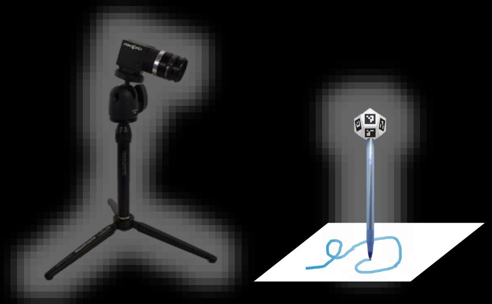
Low-cost

Contributions

- OPT Dataset
  - We build a large-scale object pose tracking benchmark dataset consisting of RGB-D video sequences of 2D and 3D targets.

- DPE Algorithm
  - We propose a robust direct approach of 6DoF pose estimation for planar objects.

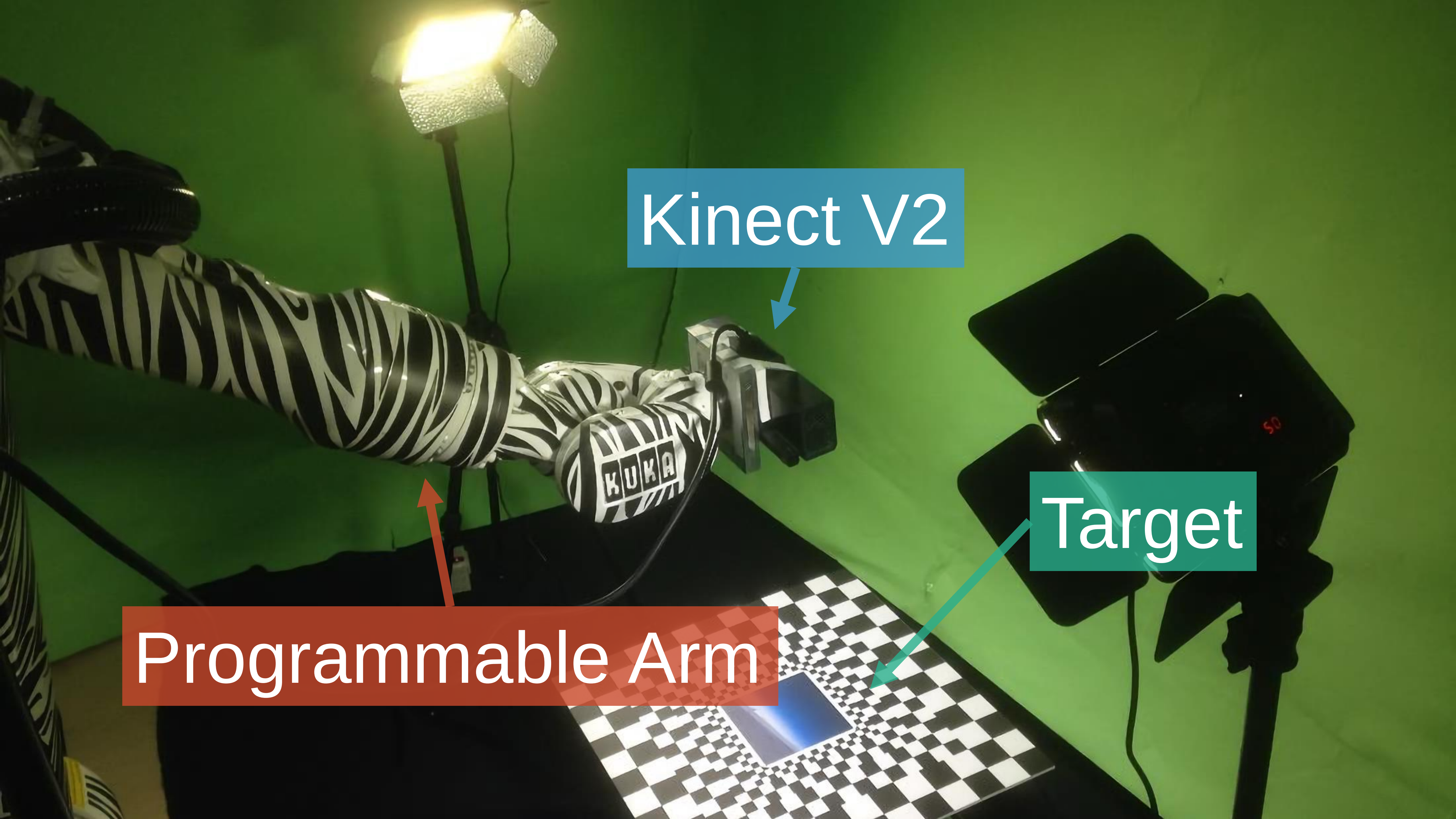
- DodecaPen
  - We develop a solution for real-time 6DoF tracking that achieves submillimeter accuracy.



Object Pose Tracking

OPT

Dataset


Kinect V2
KUKA
Target
Programmable Arm

2D Objects

Low Texture	Normal Texture	High Texture

ProJet® 460Plus

https://www.3dsystems.com/3d-printers/projet-cjp-460plus

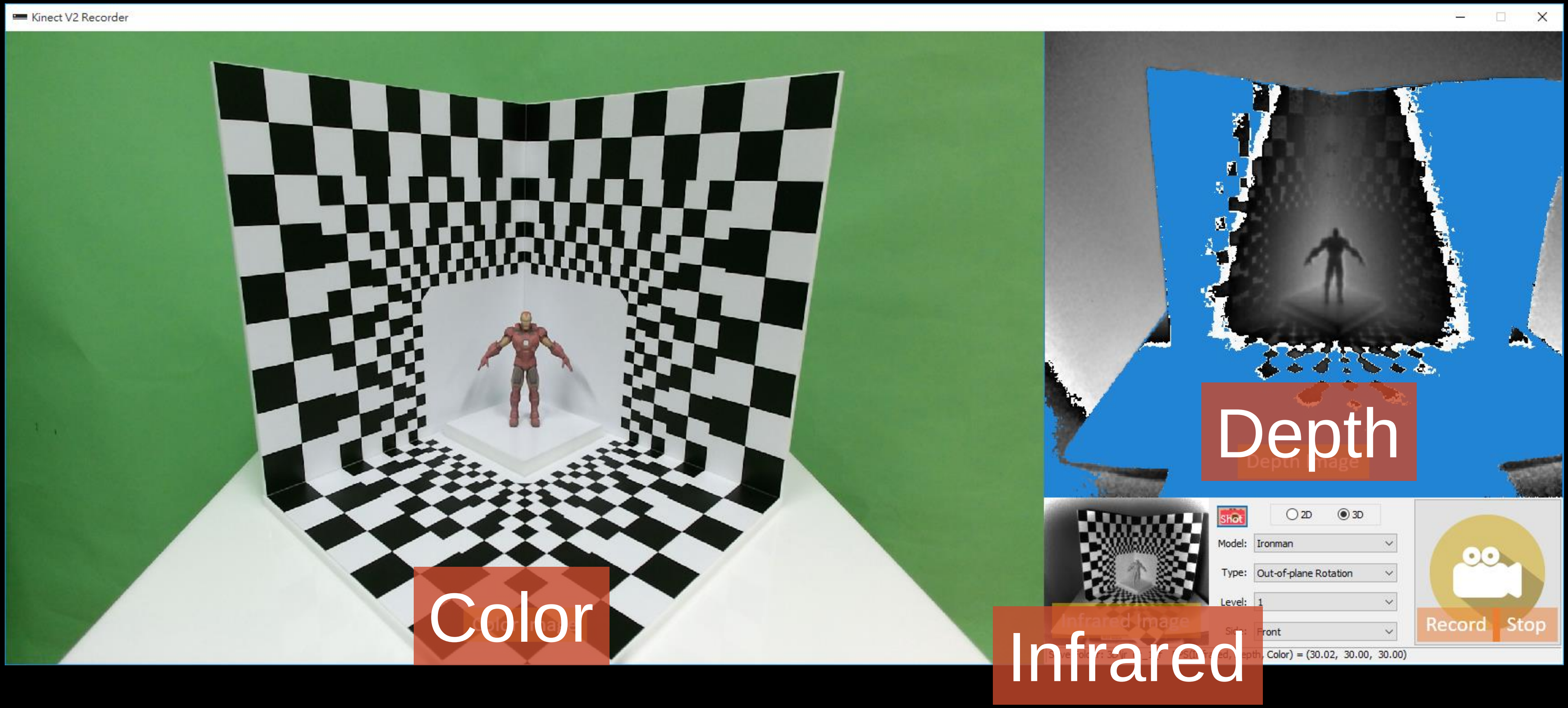
3D Objects

Simple Geometry		Normal Geometry		Complex Geometry	
					
					

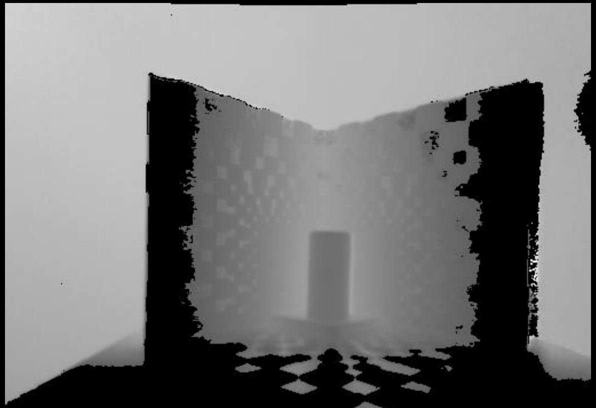
http://www.arzapstudio.com/wp-content/uploads/2016/12/Kinect-Banner-800x321.jpg

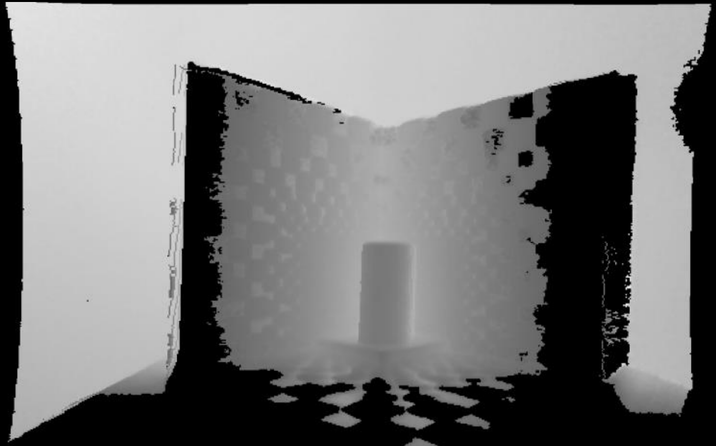
Kinect V2	RGB	Depth	Infrared
Resolution	1920x1080	512x424	512x424
Shutter Type	Rolling	Global	Global

KinectV2 Recorder



Relative Rigid Transformation





http://img.directindustry.com/images_di/photo-mg/17587-2781073.jpg

Robotic Arm: KUKA KR 16-2 CR

- Payload: 16 kg
- Repeatability: 0.05 mm
- Max. reach: 1610 mm
- Number of axes: 6

7 Motion Patterns

with 5 Speed Levels

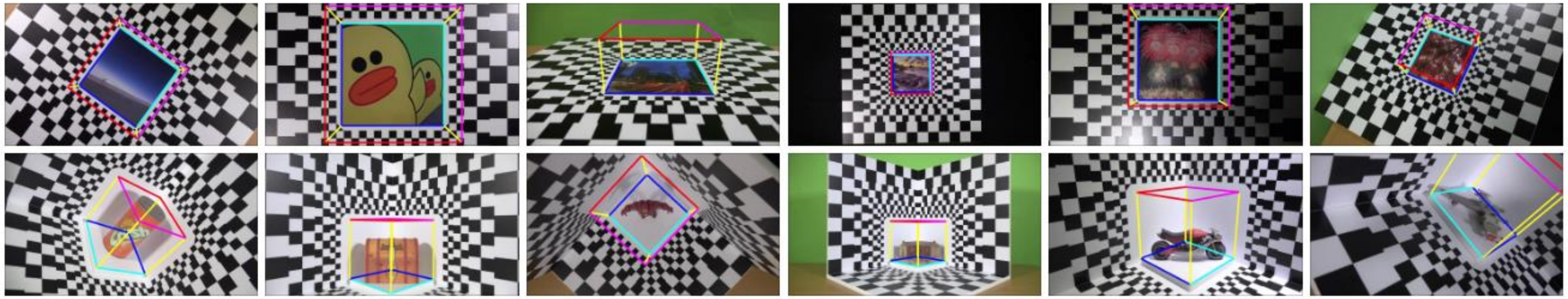
Take

Home Abstract Paper Download Results References

A Benchmark Dataset for 6DoF Object Pose Tracking

Po-Chen Wu[1] Yueh-Ying Lee[1*] Hung-Yu Tseng[1*] Hsuan-I Ho[1*] Ming-Hsuan Yang[2] Shao-Yi Chien[1]

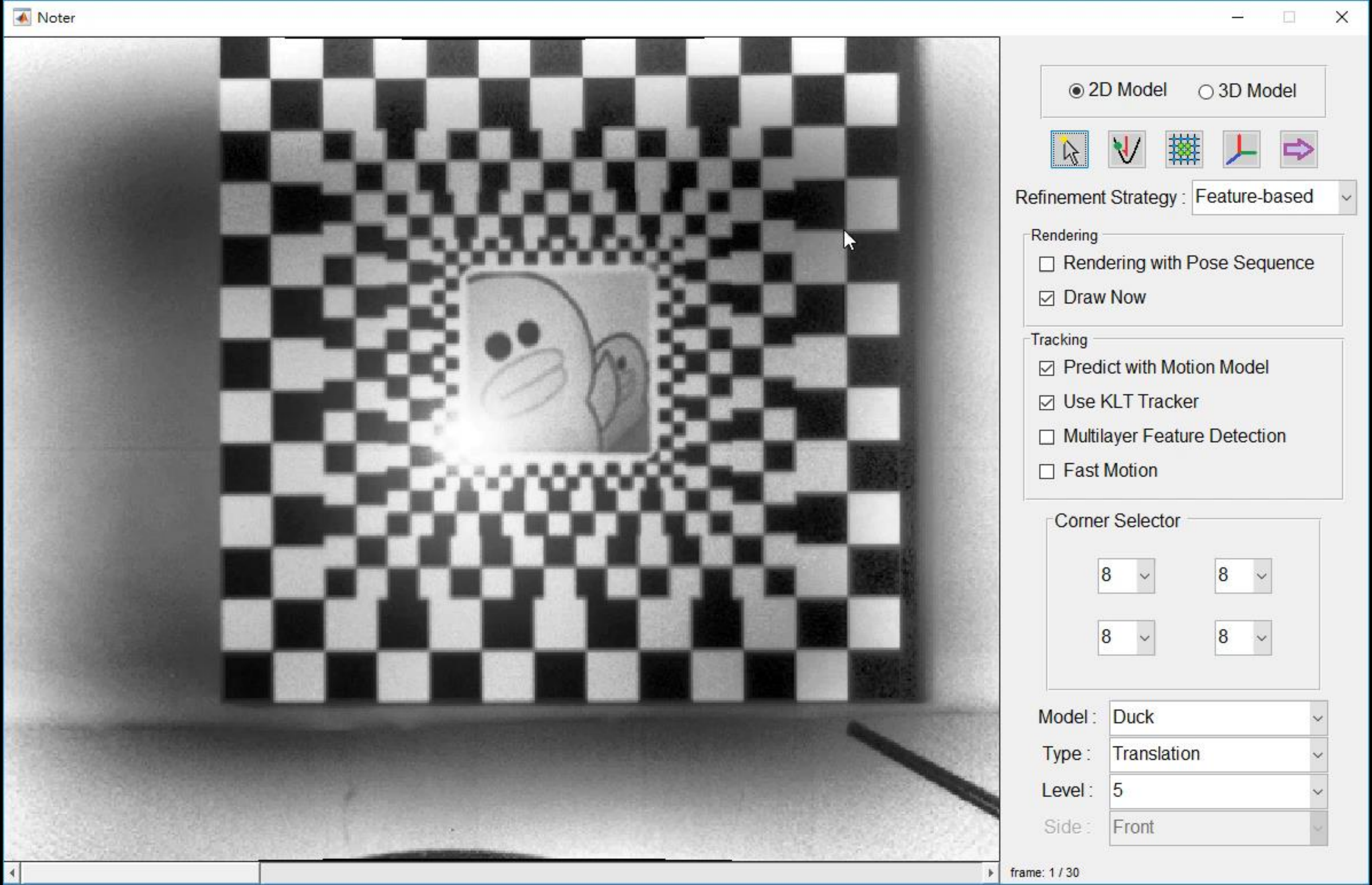
[1] Media IC & System Lab, National Taiwan University
[2] Vision and Learning Lab, University of California, Merced
(* indicates equal contribution)

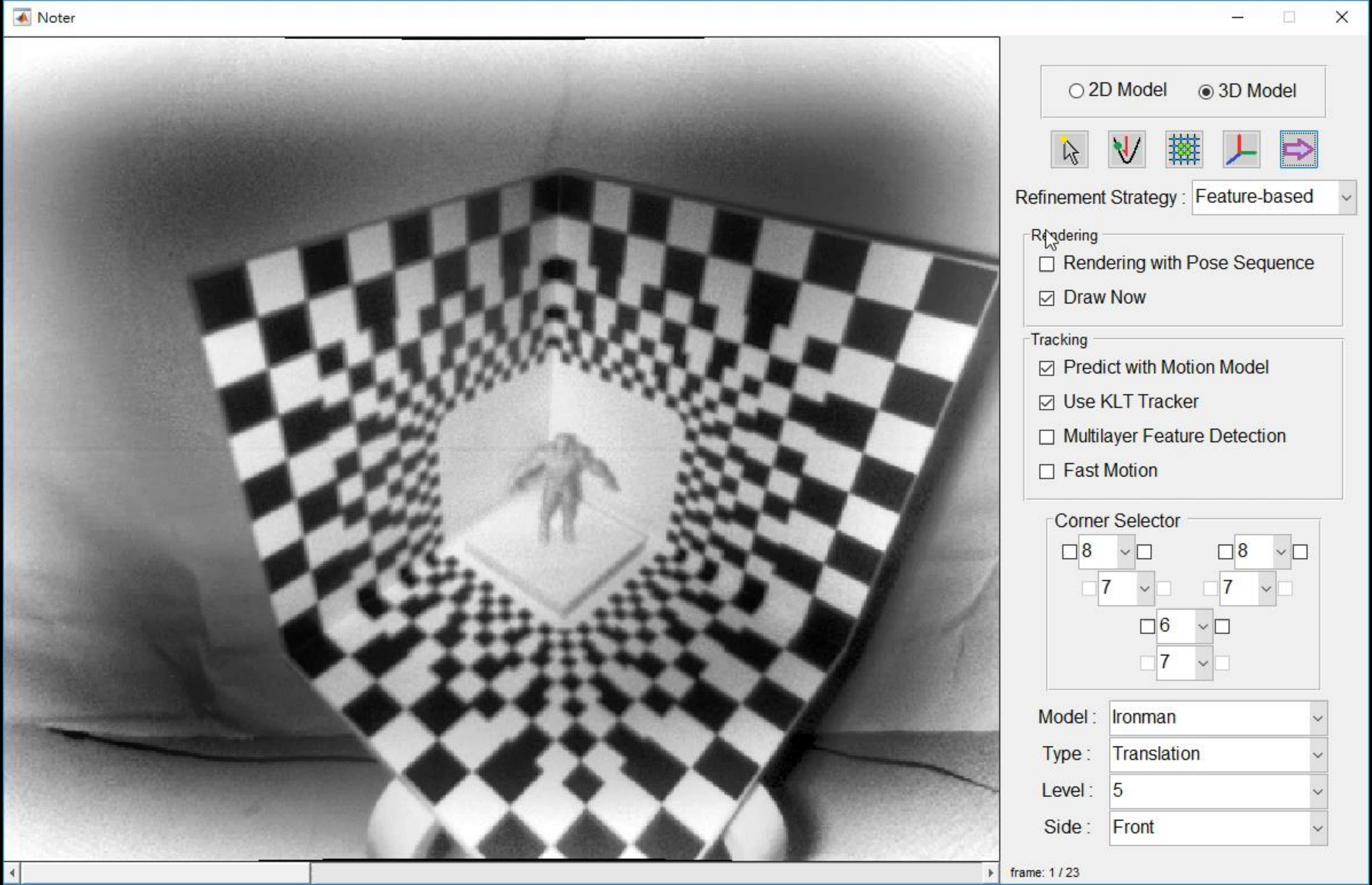


Abstract

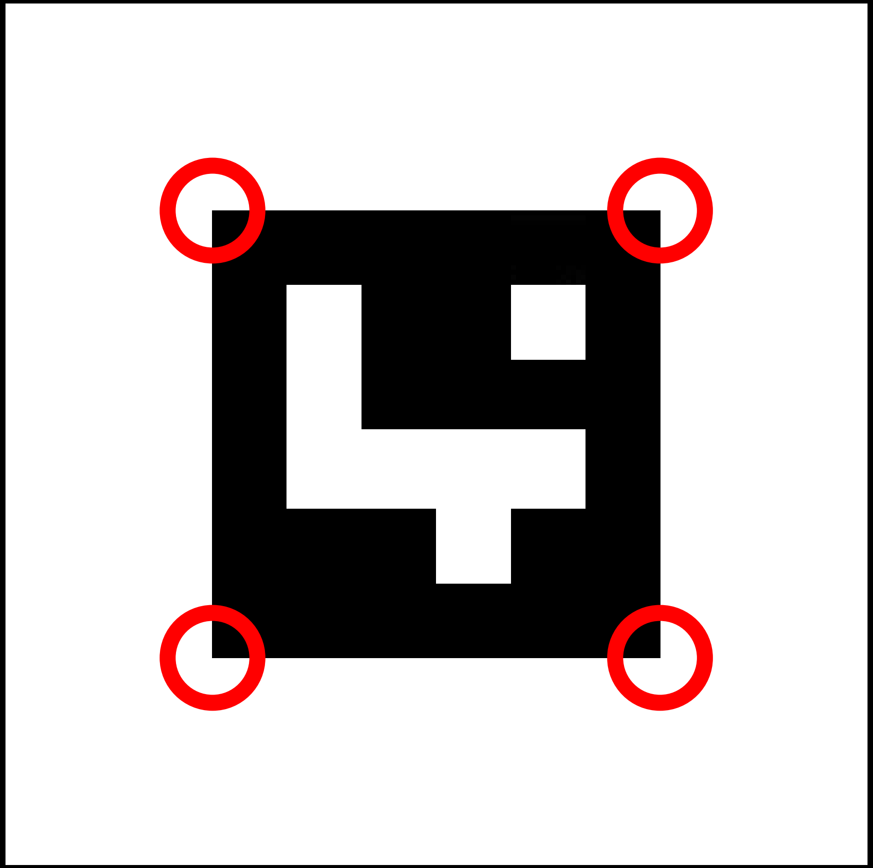
Accurately tracking the six degree-of-freedom pose of an object in real scenes is an important task in computer vision and augmented reality with numerous applications. Although a variety of algorithms for this task have been proposed, it remains difficult to evaluate existing methods in the literature as oftentimes different [illegible] datasets close to [illegible] are available. In this paper, we present a large object pose tracking benchmark dataset consisting of RGB-D video sequences of 2D and 3D targets with ground-truth information. The videos are recorded under various lighting conditions, different motion patterns and speeds with the help of a programmable robotic arm. We present extensive quantitative evaluation results of the state-of-the-art methods on this benchmark dataset and discuss the potential research directions in this field.

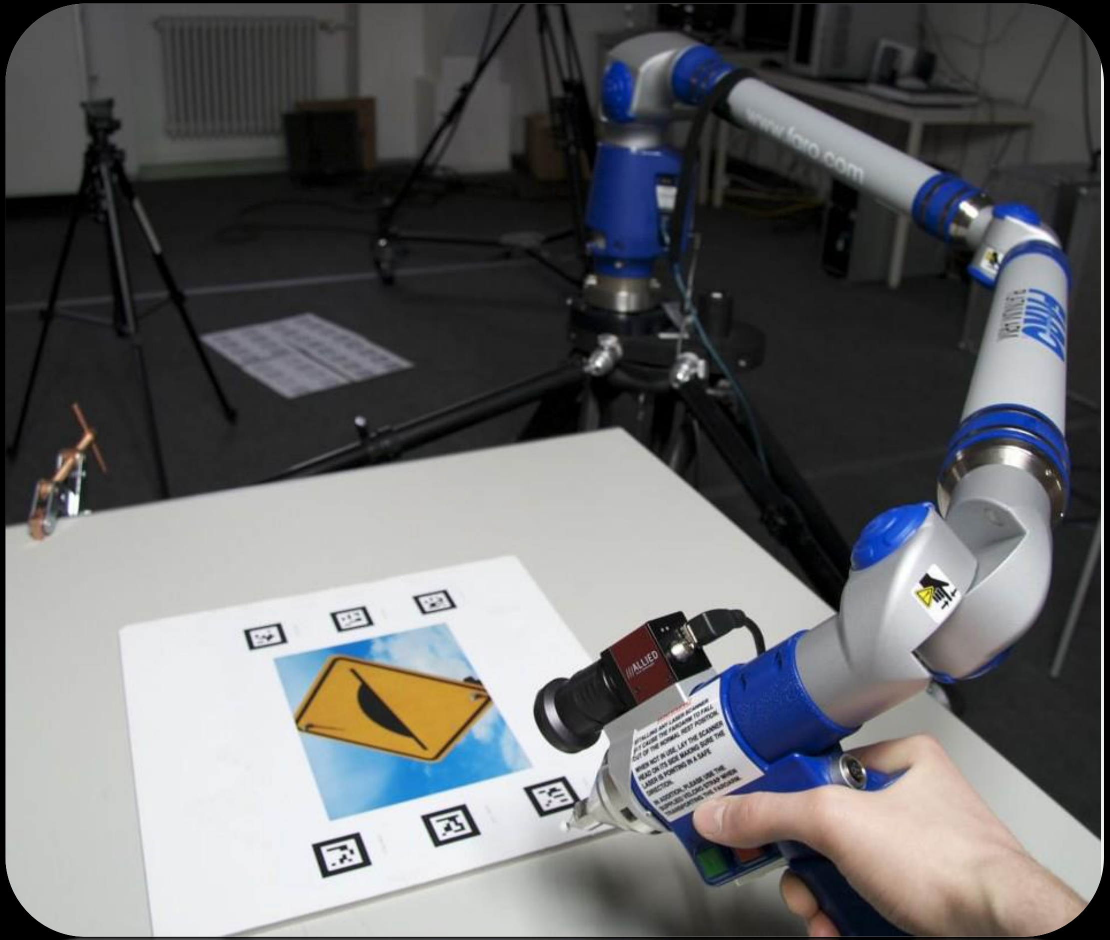
http://media.ee.ntu.edu.tw/research/OPT/


Noter
2D Model
3D Model
Refinement Strategy : Feature-based
Rendering
Rendering with Pose Sequence
Draw Now
Tracking
Predict with Motion Model
Use KLT Tracker
Multilayer Feature Detection
Fast Motion
Corner Selector
8
8
8
8
Model : Duck
Type : Translation
Level : 5
Side : Front
frame: 1 / 30

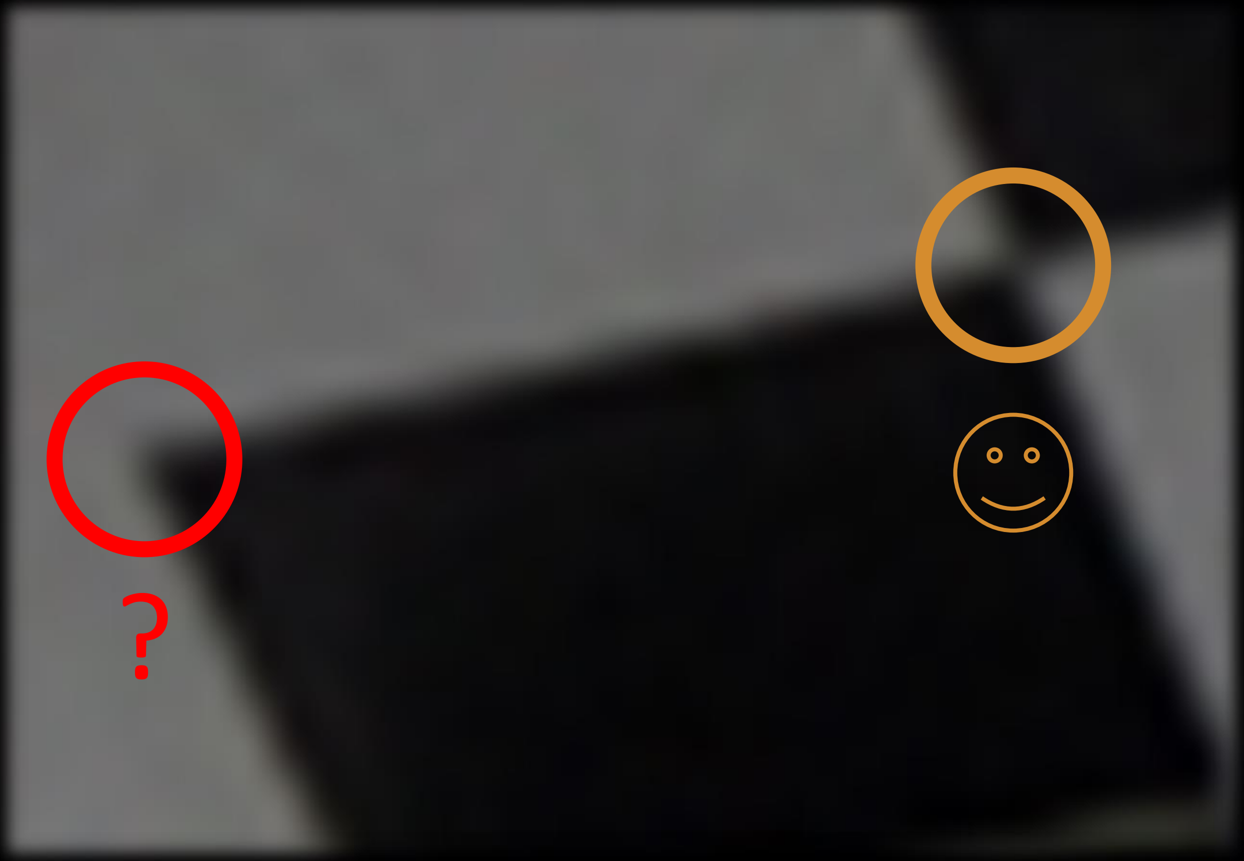

Noter
2D Model
3D Model
Refinement Strategy :
Feature-based
Rendering
Rendering with Pose Sequence
Draw Now
Tracking
Predict with Motion Model
Use KLT Tracker
Multilayer Feature Detection
Fast Motion
Corner Selector
8
8
7
7
6
7
Model :
Ironman
Type :
Translation
Level :
5
Side :
Front
frame: 1 / 23

Marker Corner Localization

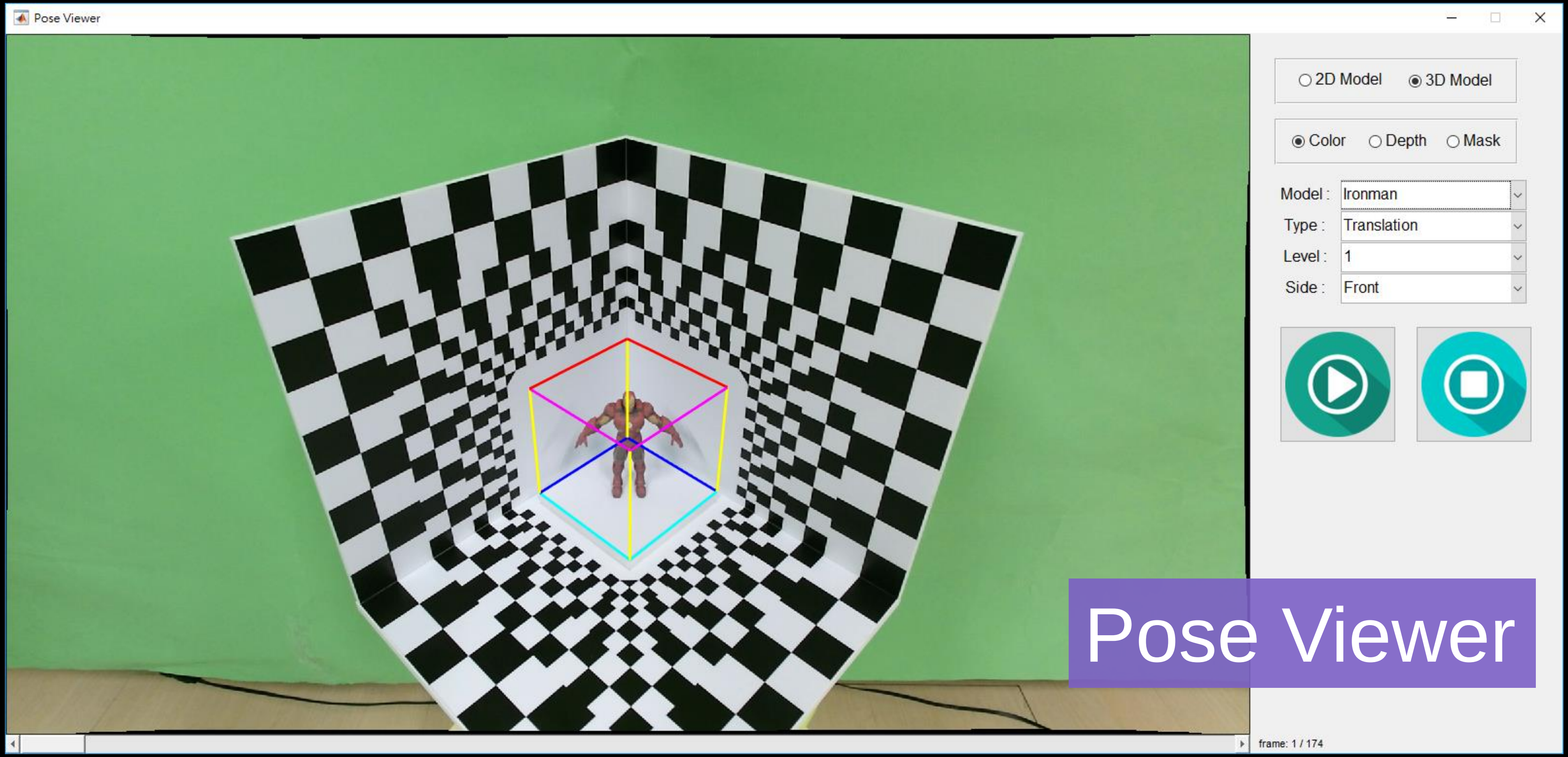


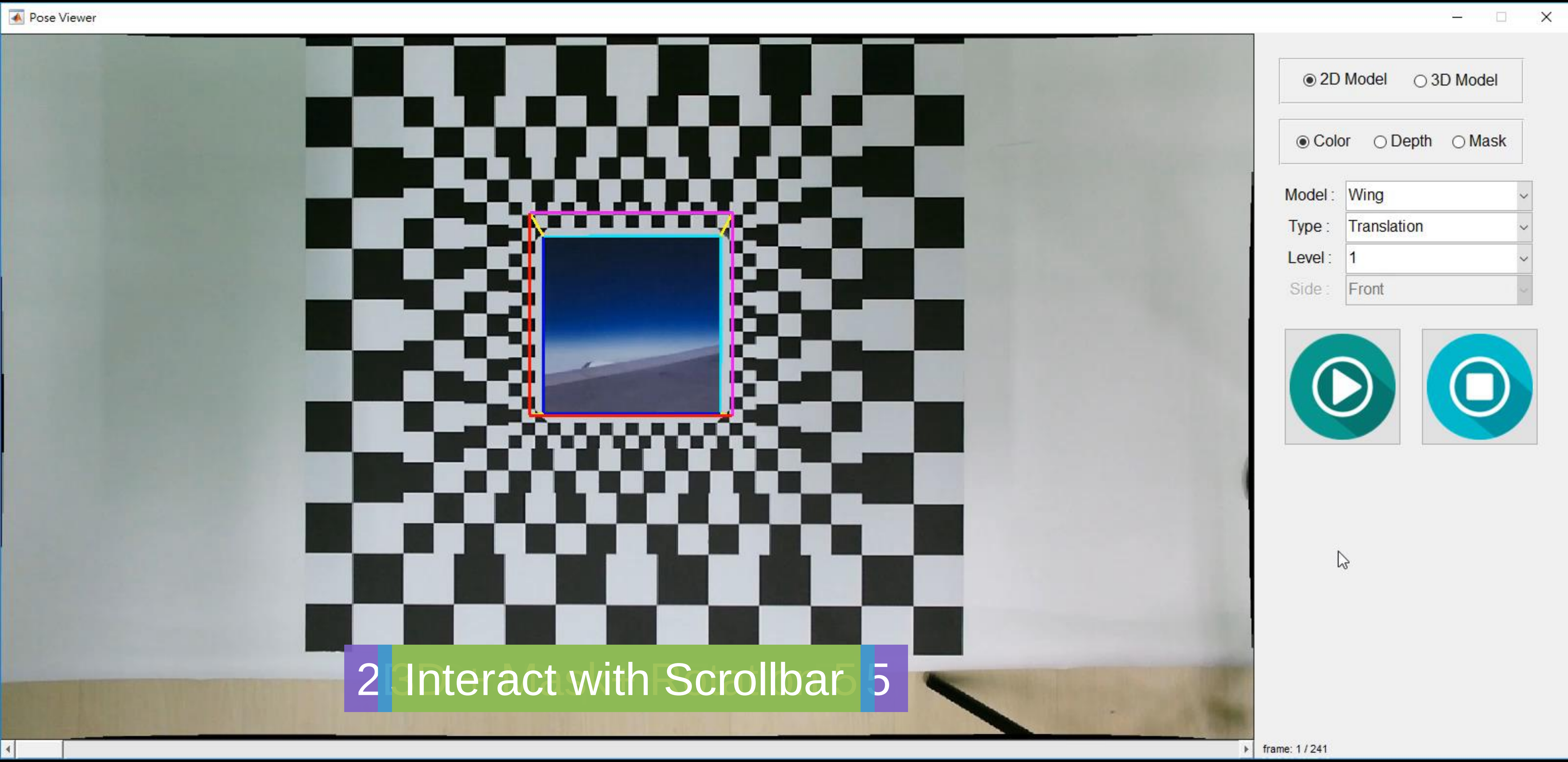


Corner Localization Accuracy

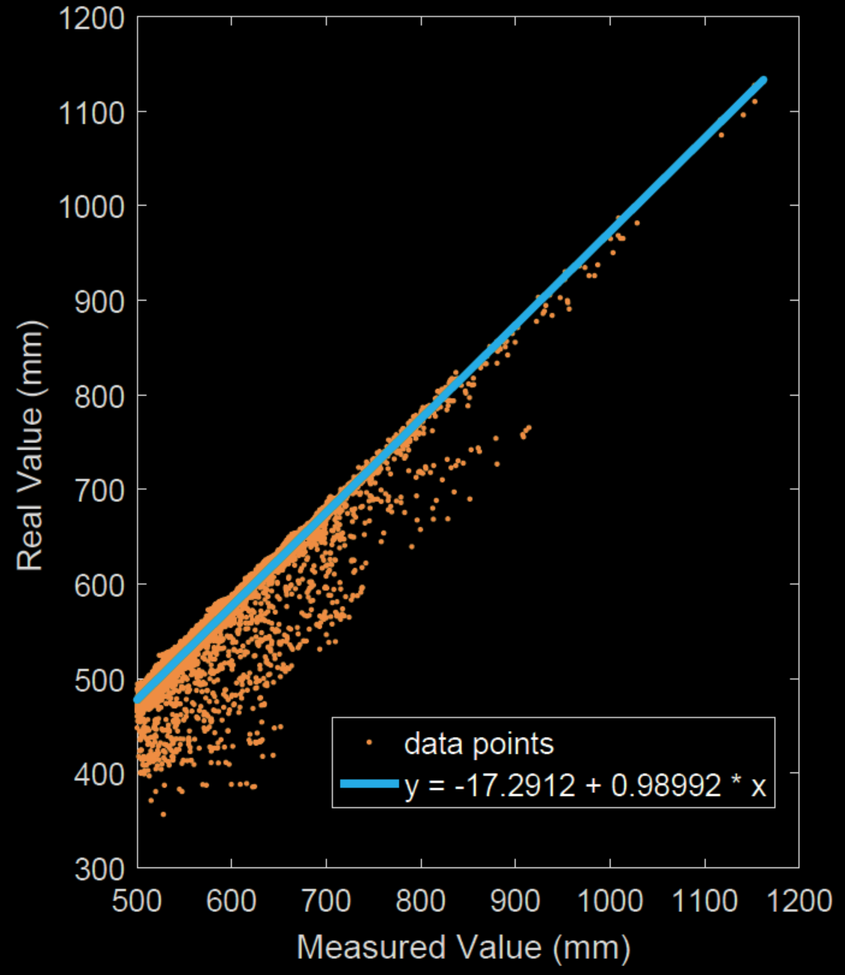


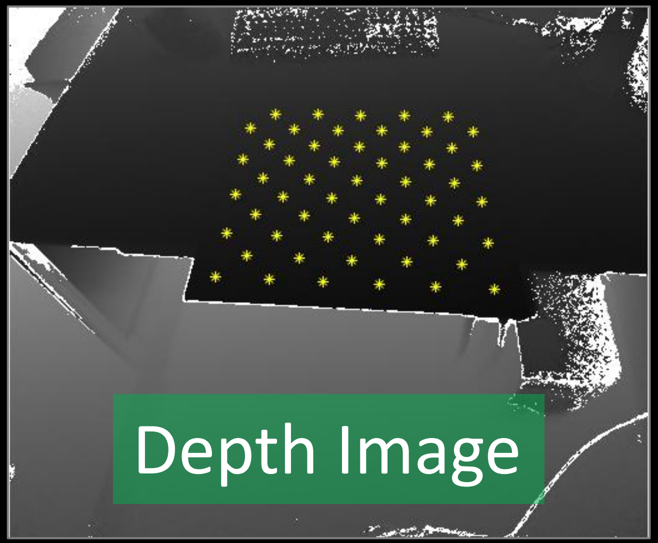
Ground-truth Pose

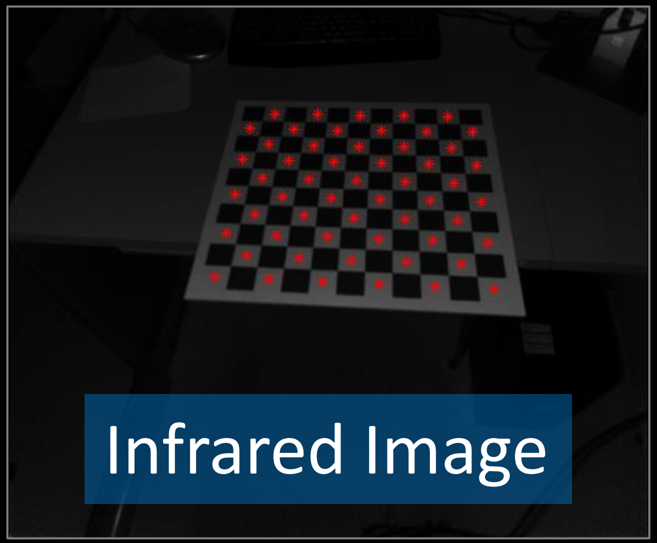



Pose Viewer
2D Model
3D Model
Color
Depth
Mask
Model : Wing
Type : Translation
Level : 1
Side : Front
2 Interact with Scrollbar 5
frame: 1 / 241

Depth Calibration

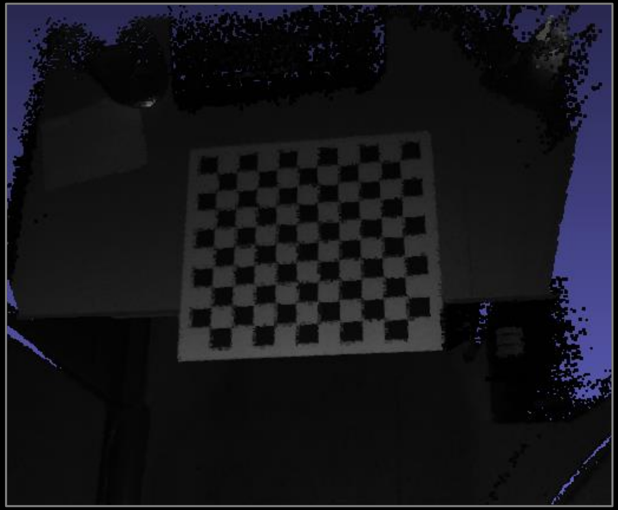




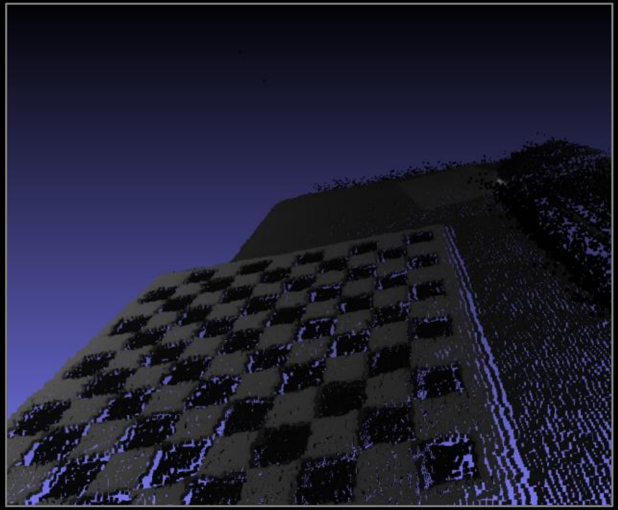


Measured Depth Data

Infrared Image

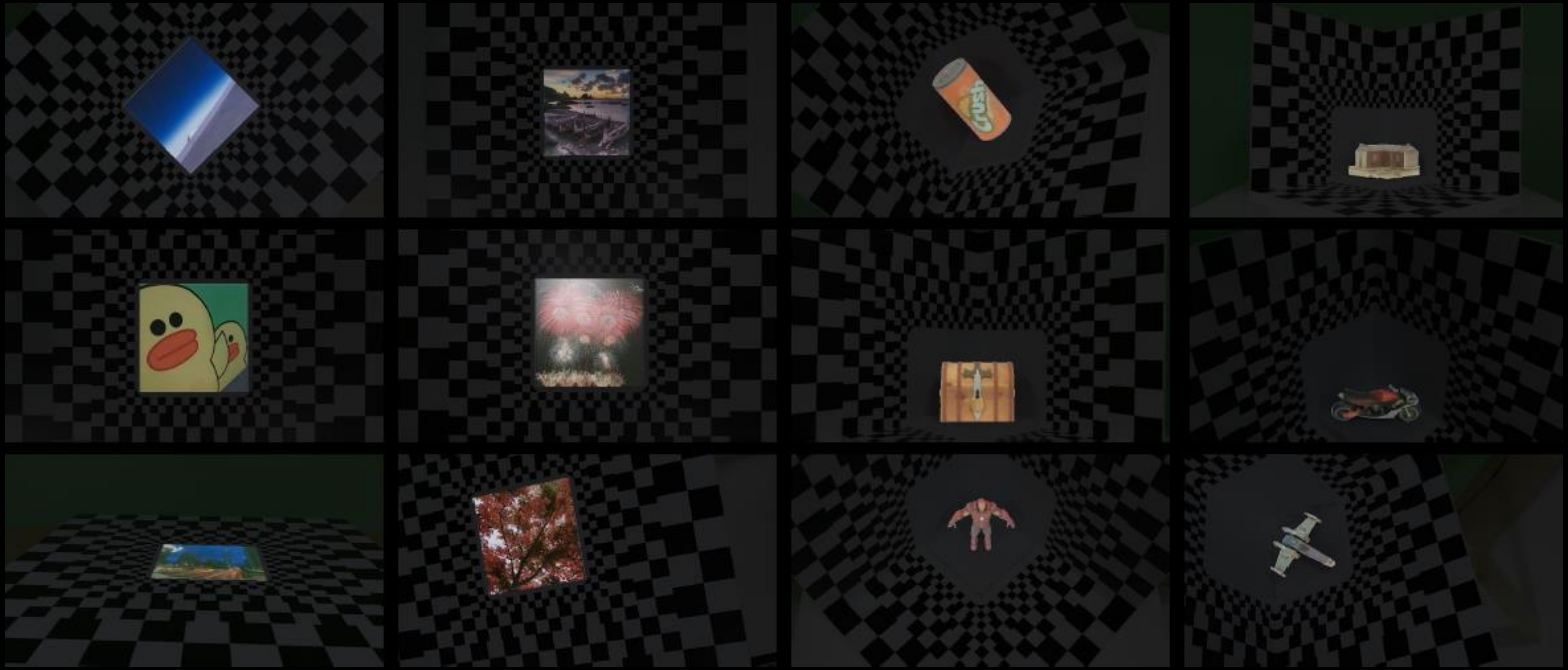


Point Cloud (a)



Point Cloud (b)

Masked Image

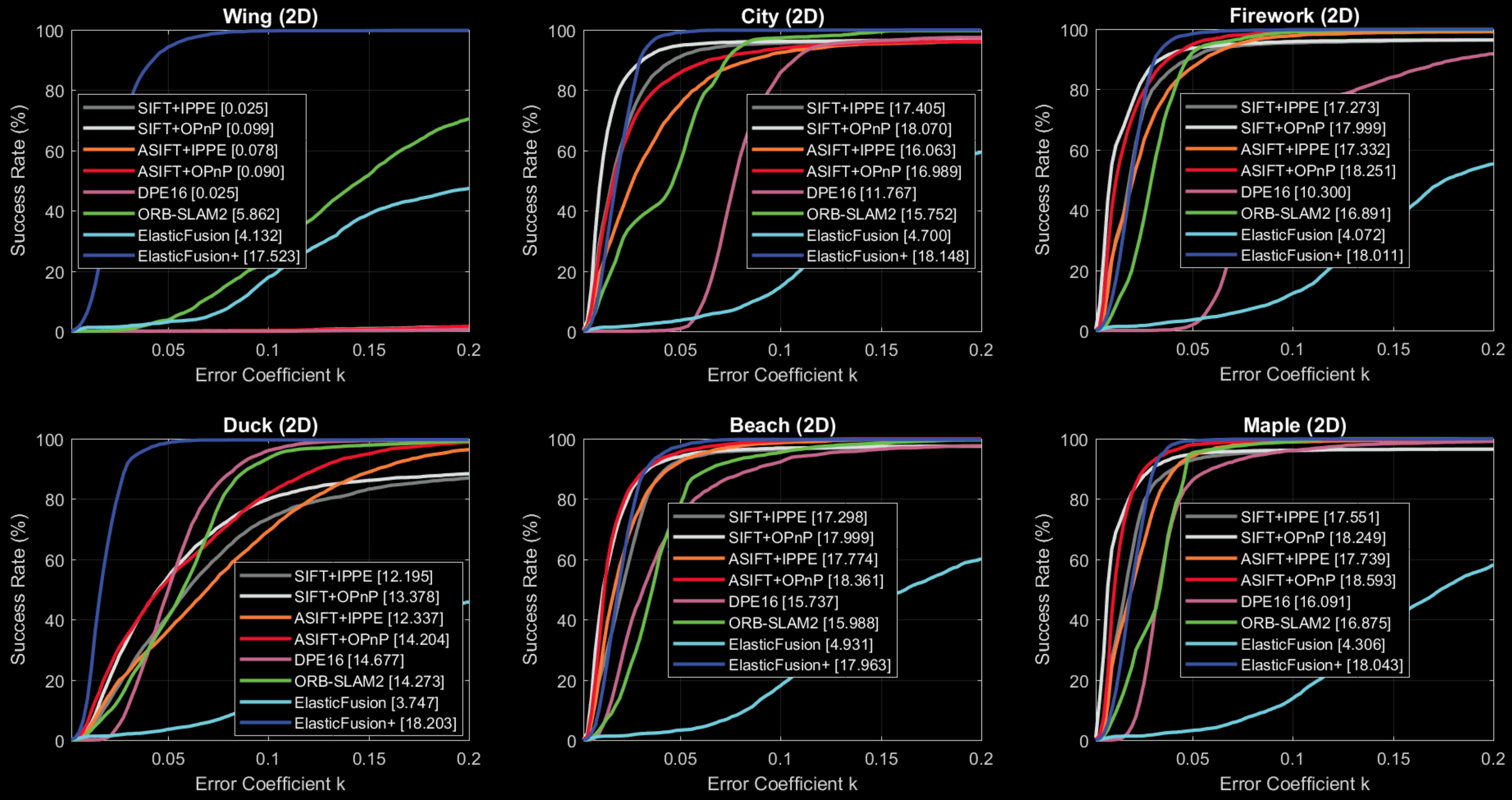


Dataset Comparison

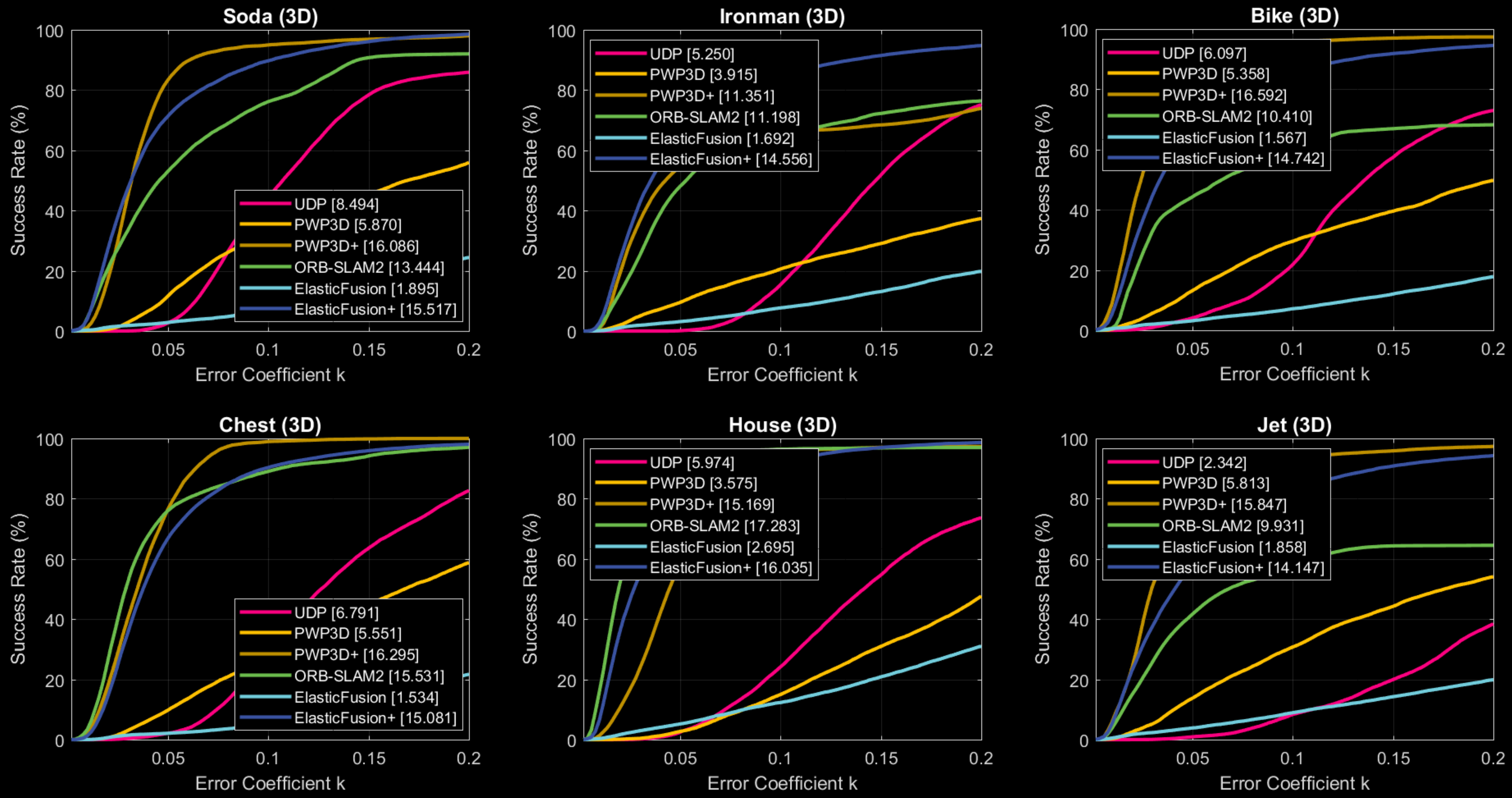
Benchmark	Device	Mechanism	Pose Establishment	Video Clips	# 2D Targets	# 3D Targets	# Motion Patterns	# Frames
Lieberknecht [1]	Marlin F-080C	Handheld	Marker-based	Yes	8	--	5	48,000
Gauglitz [2]	Fire-i	Manually Operated Contraption	Direct Alignment	Yes	6	--	16	6,889
Hinterstoisser [3]	Kinect V1	Handheld	Marker-based	No	--	15	--	18,000
Tejani [4]	Kinect V1	Handheld	Marker-based	No	--	3	--	5,229
Brachmann [5]	Kinect V1	Handheld	Marker-based	No	--	20	3	10,000
Rennie [6]	Kinect V1	Robotic Arm	Manual	No	--	24	--	10,368
Krull [7]	Kinect V1	Handheld	ICP	Yes	--	3	--	1,000
Choi [8]	Synthetic	--	Synthetic	Yes	--	4	--	4,000
Proposed	**Kinect V2**	**Programmable Robotic Arm**	**Checkerboard-based**	Yes	6	3	23	100,956

1. S. Lieberknecht, S. Benhimane, P. Meier, and N. Navab. A Dataset and Evaluation Methodology for Template-based Tracking. In ISMAR, 2009.
2. S. Gauglitz, T. H¨ollerer, and M. Turk. Evaluation of Interest Point Detectors and Feature Descriptors for Visual Tracking. IJCV, 94(3):335– 360, 2011.
3. S. Hinterstoisser, V. Lepetit, S. Ilic, S. Holzer, G. Bradski, K. Konolige, and N. Navab. Model Based Training, Detection and Pose Estimation of Texture-Less 3D Objects in Heavily Cluttered Scenes. In ACCV, 2012.
4. A. Tejani, D. Tang, R. Kouskouridas, and T.-K. Kim. Latent-Class Hough Forests for Object Detection and Pose Estimation. In ECCV, 2014.
5. E. Brachmann, A. Krull, F. Michel, S. Gumhold, J. Shotton, and C. Rother. Learning 6D Object Pose Estimation Using 3D Object Coordinates. In ECCV, 2014.
6. C. Rennie, R. Shome, K. E. Bekris, and A. F. De Souza. A Dataset for Improved RGBD-based Object Detection and Pose Estimation for Warehouse Pick-and-Place. RAL, 1(2):1179–1185, 2016.
7. A. Krull, F. Michel, E. Brachmann, S. Gumhold, S. Ihrke, and C. Rother. 6-DOF Model Based Tracking via Object Coordinate Regression. In ACCV, 2014.
8. C. Choi and H. I. Christensen. RGB-D Object Tracking: A Particle Filter Approach on GPU. In IROS, 2013.

Evaluation on 2D Datasets

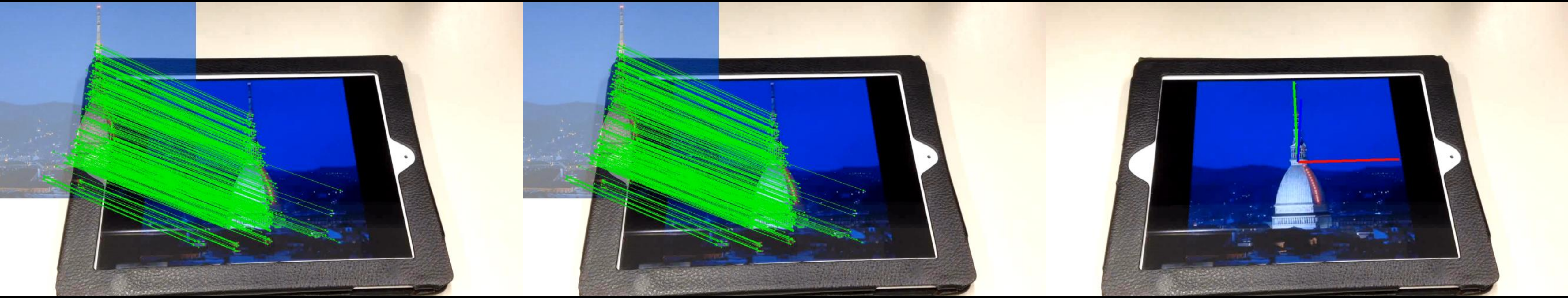


Evaluation on 3D Datasets



Feature-Based Method

Feature Matching → Outlier Removal → PnP Algorithm

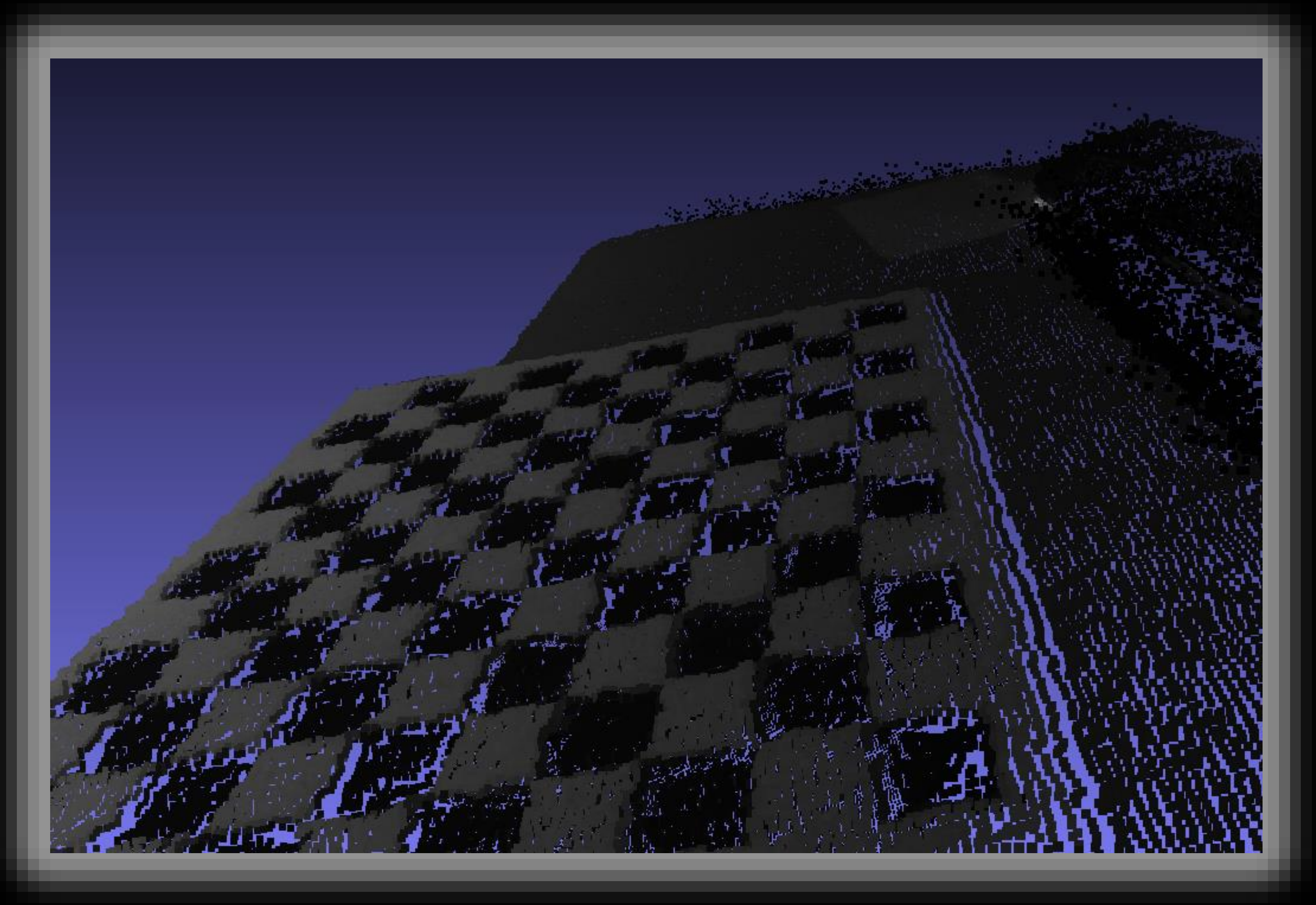


Rely on Natural Features

Would Fail When...

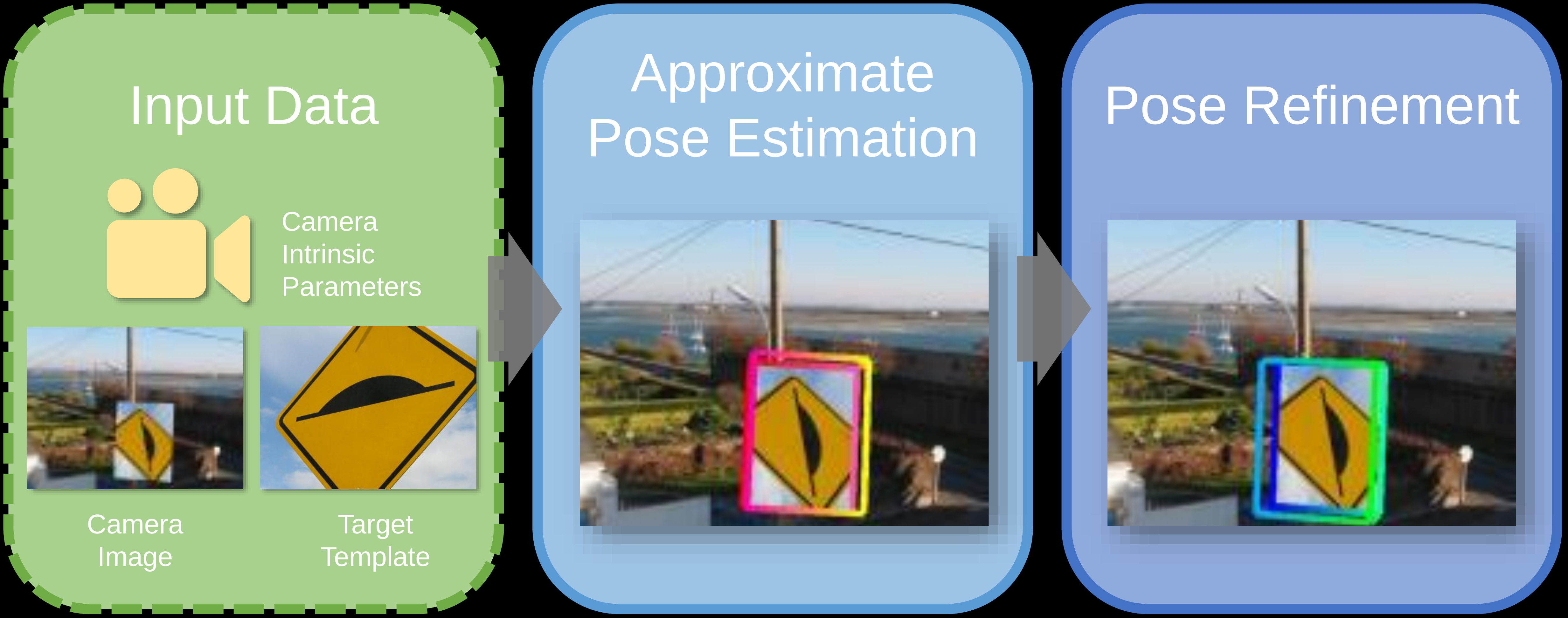
Textureless

Blurry



Noisy Depth Data

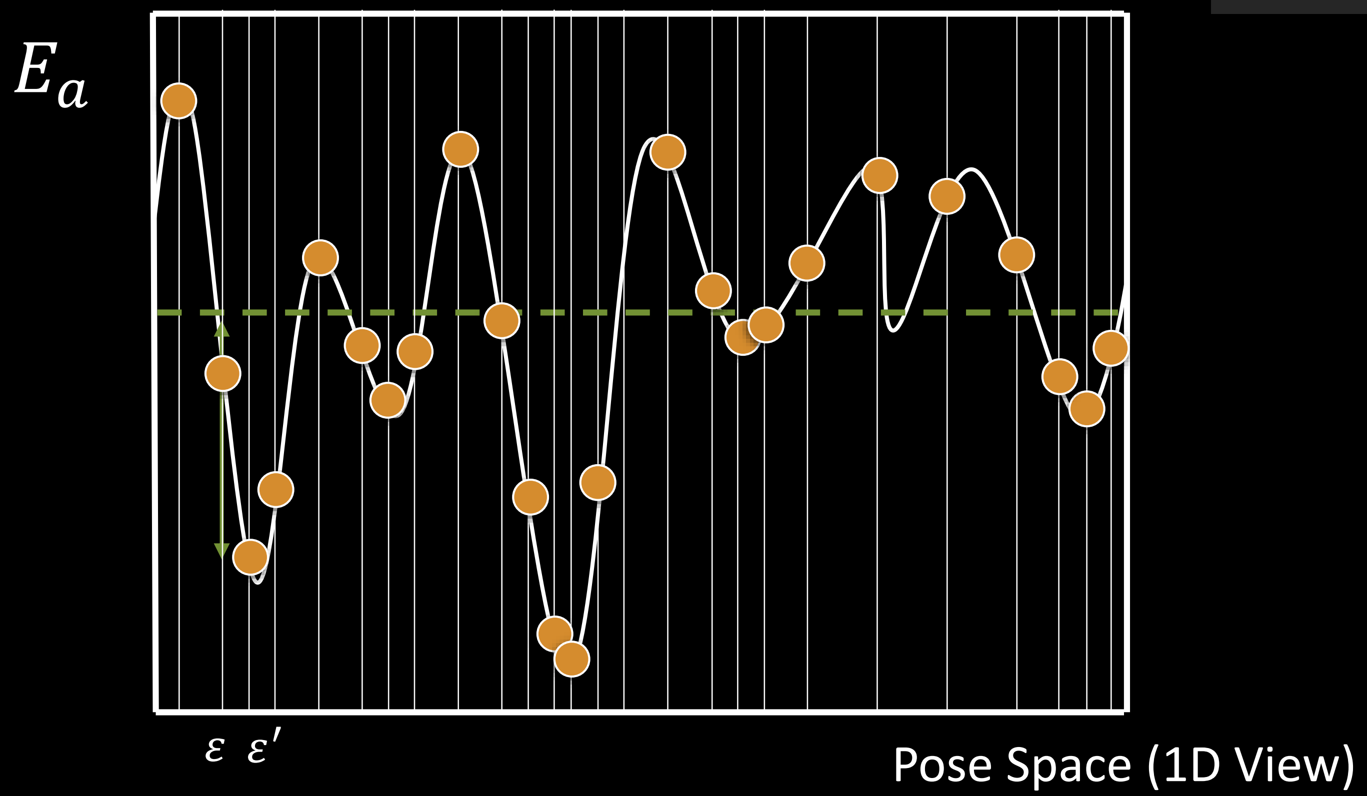
Direct Pose Estimation (DPE)



Approximate Pose Estimation (APE)

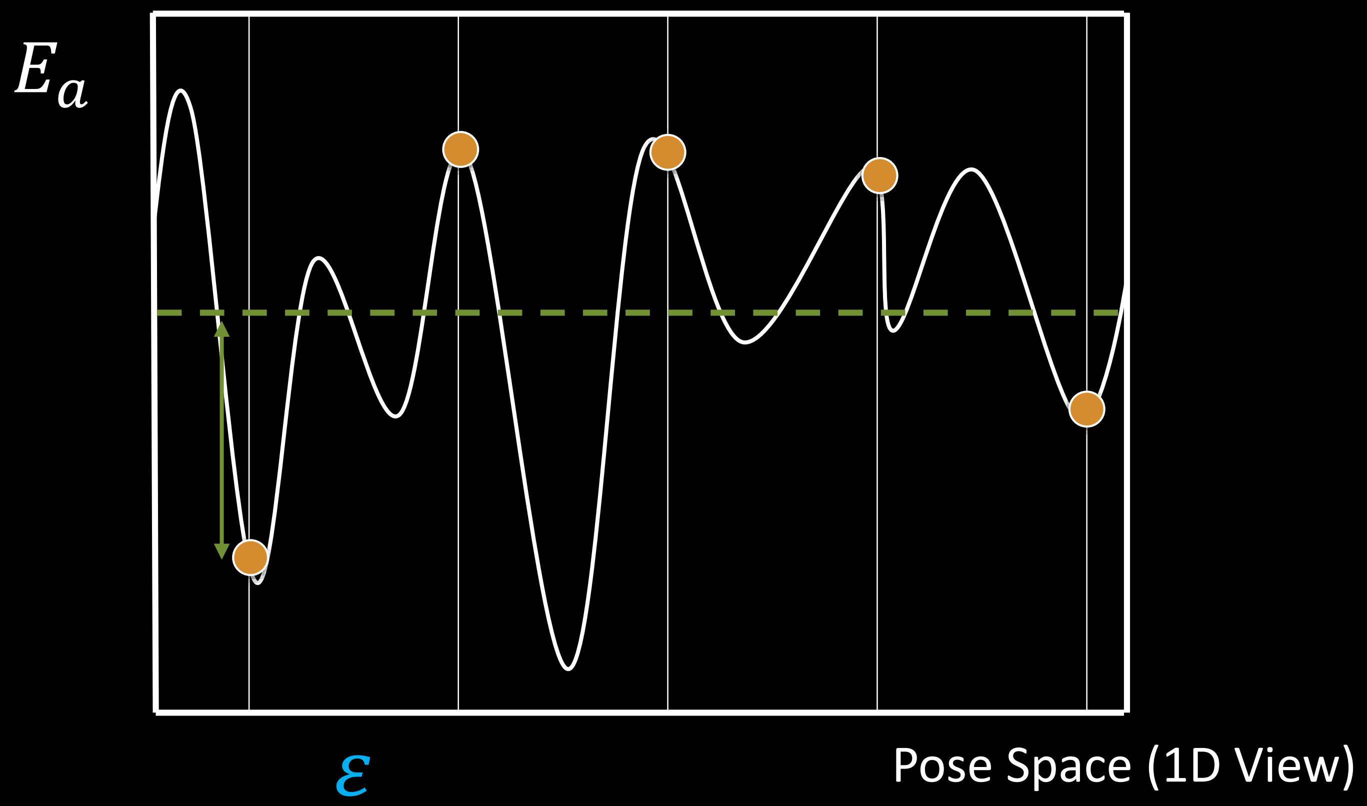
- Branch-and-bound Algorithm in Pose Space
  - Find the pose with minimum appearance error E_a

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

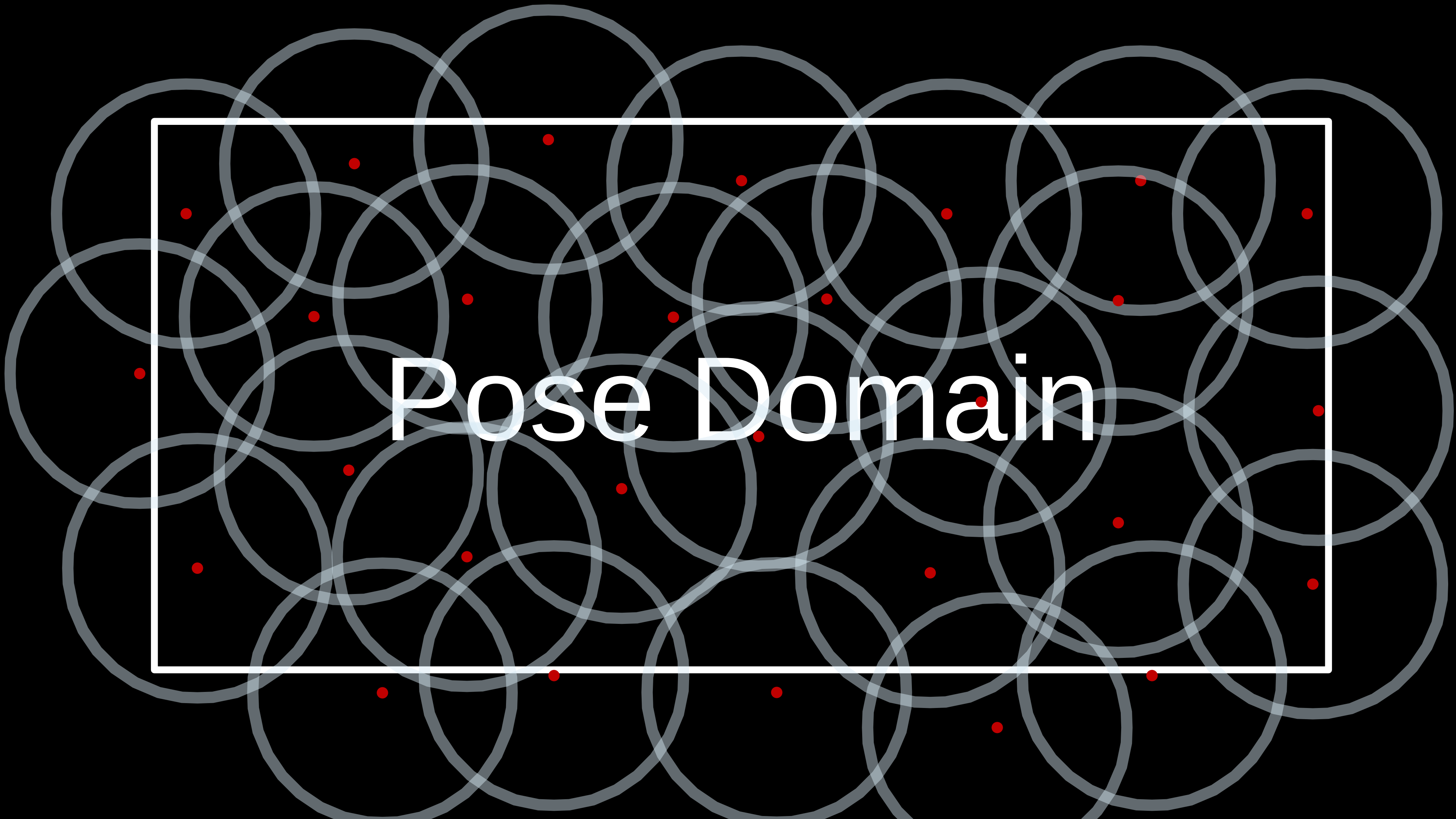


Approximate Pose Estimation (APE)

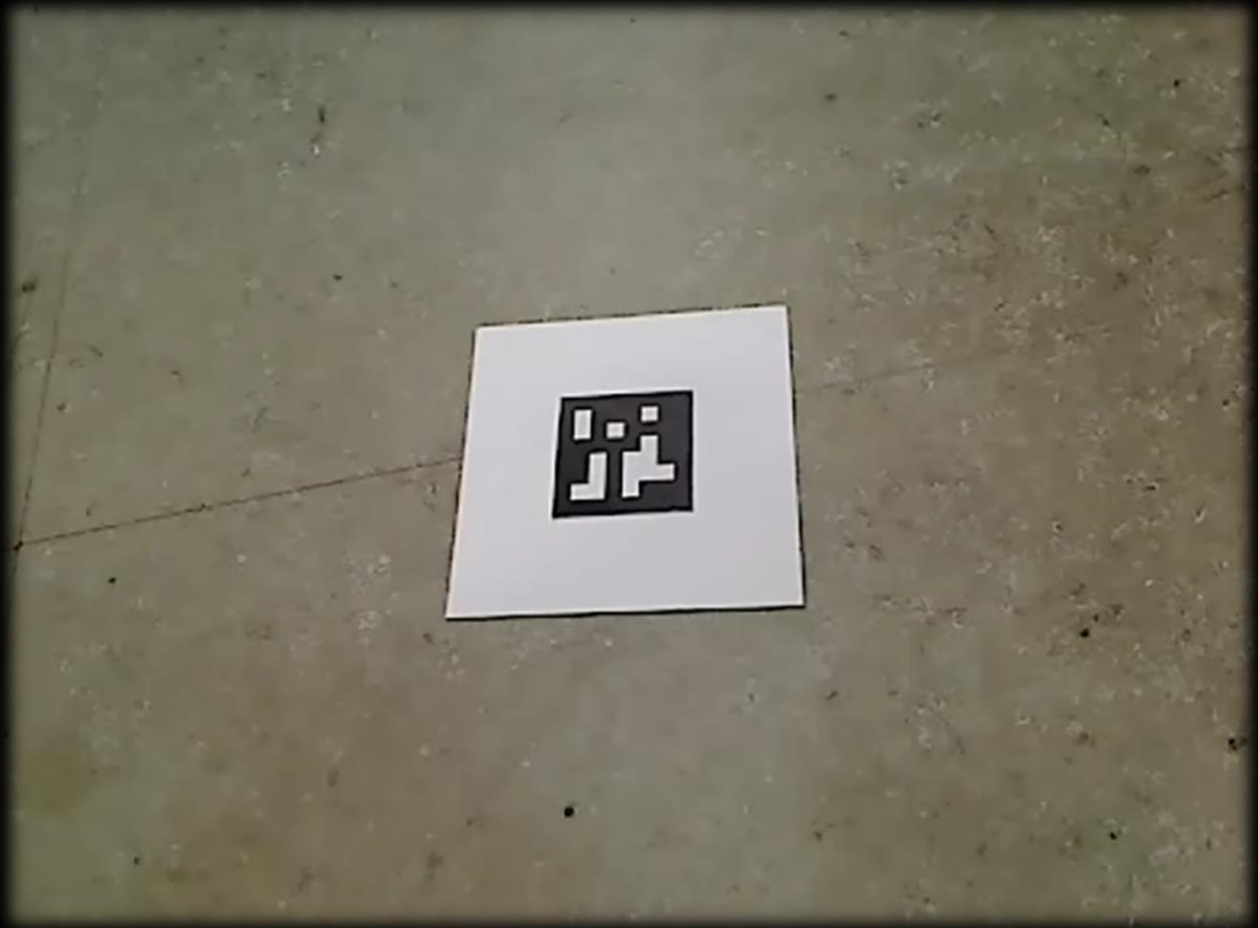
- Failure analysis



Ɛ-covering Set



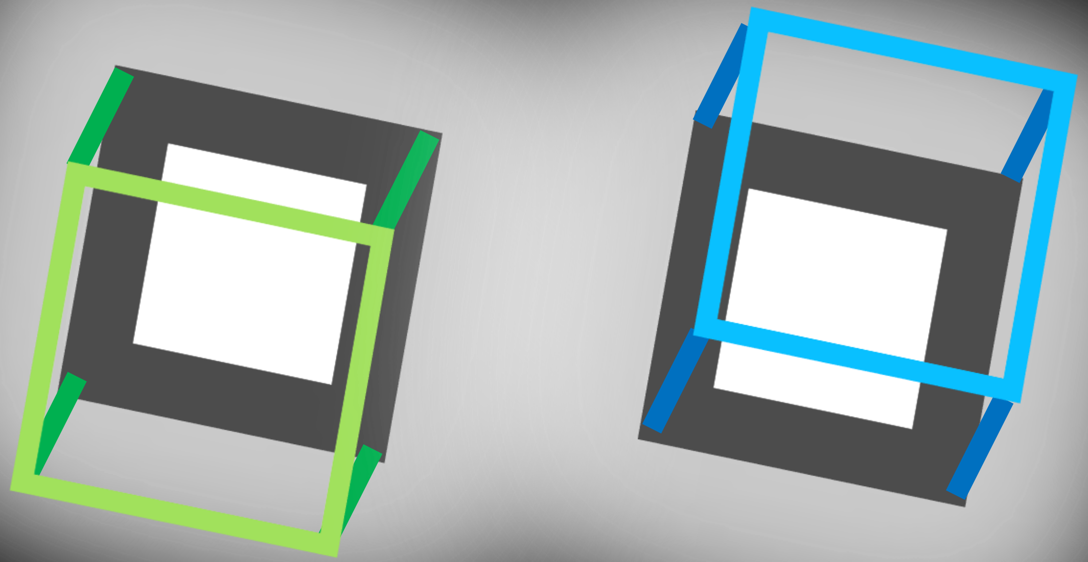
Pose Jumping



Original Video

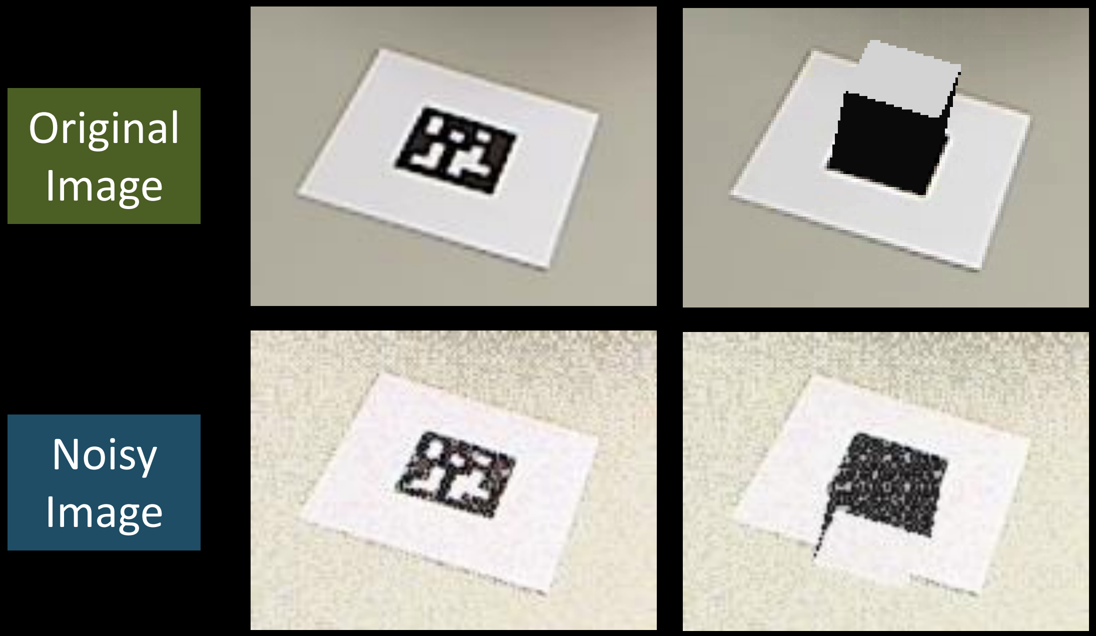
Pose Estimation Result

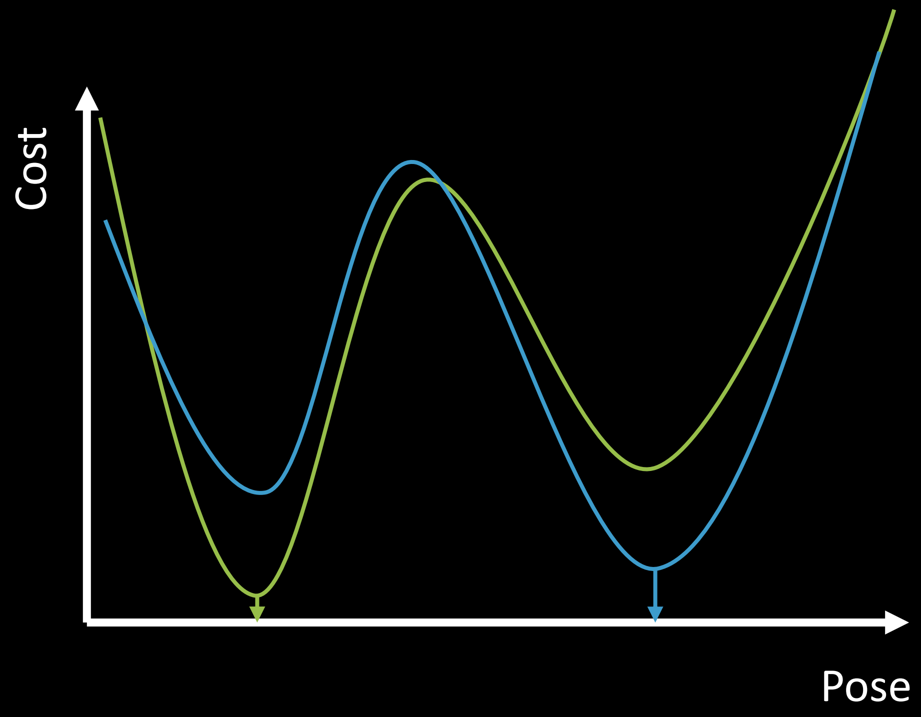
Pose Ambiguity



Explanation of Pose Ambiguity

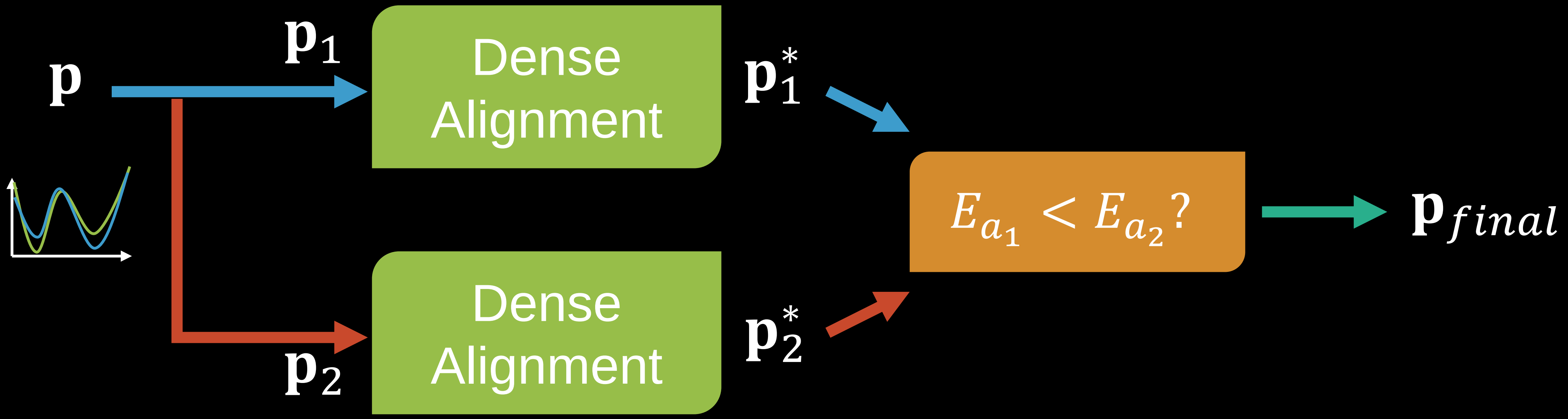
- Multiple local minima of a cost function (due to coplanar points)





Pose Refinement (PR)

- Refine and disambiguate the approximately estimated pose



Gauss-Newton Iteration

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

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

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

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

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

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

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

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

- Chain rule

$$\mathbf{J}_c = \frac{\partial \mathbf{I}_c}{\partial \mathbf{p}} = \begin{bmatrix} \frac{\partial I_c(\mathbf{u}_1)}{\partial \mathbf{p}} \\ \vdots \\ \frac{\partial I_c(\mathbf{u}_n)}{\partial \mathbf{p}} \end{bmatrix}, \frac{\partial I_c}{\partial \mathbf{p}} = \frac{\partial I_c}{\partial \mathbf{u}} \left[\frac{\partial \mathbf{u}}{\partial \mathbf{r}}, \frac{\partial \mathbf{u}}{\partial \mathbf{t}} \right] = \left[\frac{\partial I_c}{\partial u}, \frac{\partial I_c}{\partial v} \right] \left[\frac{\partial \mathbf{u}}{\partial \hat{\mathbf{x}}} \frac{\partial \hat{\mathbf{x}}}{\partial \hat{\mathbf{R}}} \frac{\partial \hat{\mathbf{R}}}{\partial \mathbf{r}}, \frac{\partial \mathbf{u}}{\partial \hat{\mathbf{x}}} \right],$$

$$\frac{\partial \mathbf{u}}{\partial \hat{\mathbf{x}}} = \begin{bmatrix} \frac{f_x}{\hat{z}} & 0 & -\frac{f_x \hat{x}}{\hat{z}^2} \\ 0 & \frac{f_y}{\hat{z}} & -\frac{f_y \hat{y}}{\hat{z}^2} \end{bmatrix}, \frac{\partial \hat{\mathbf{x}}}{\partial \hat{\mathbf{R}}} = \begin{bmatrix} x & y & 0 & 0 & 0 & 0 \\ 0 & 0 & x & y & 0 & 0 \\ 0 & 0 & 0 & 0 & x & y \end{bmatrix}, \partial \hat{\mathbf{R}} = \begin{bmatrix} R_{11} \\ R_{12} \\ R_{21} \\ R_{22} \\ R_{31} \\ R_{32} \end{bmatrix}, \partial \hat{\mathbf{x}} = \begin{bmatrix} \hat{x} \\ \hat{y} \\ \hat{z} \end{bmatrix} = \begin{bmatrix} R_{11} & R_{12} & t_x \\ R_{21} & R_{22} & t_y \\ R_{31} & R_{32} & t_z \end{bmatrix} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix}$$

Jacobian Matrix $\mathbf{J}_c \equiv \left[\frac{\partial I_c}{\partial u}, \frac{\partial I_c}{\partial v}\right]\left[\frac{\partial \mathbf{u}}{\partial \hat{\mathbf{x}}}\frac{\partial \hat{\mathbf{x}}}{\partial \hat{\mathbf{R}}}\frac{\partial \hat{\mathbf{R}}}{\partial \mathbf{r}}, \frac{\partial \mathbf{u}}{\partial \hat{\mathbf{x}}}\right]$

- The rotation is parameterized as rotation vector

$$\mathbf{p} = \begin{bmatrix}\mathbf{r}\\ \mathbf{t}\end{bmatrix}, \mathbf{r} = \begin{bmatrix}r_x\\ r_y\\ r_z\end{bmatrix} \in \mathbb{R}^3, \mathbf{t} = \begin{bmatrix}t_x\\ t_y\\ t_z\end{bmatrix} \in \mathbb{R}^3$$

- The derivative of $\mathbf{R}$ with respect to $\mathbf{r}$[1]:

$$\frac{\partial \mathbf{R}}{\partial r_a} = \frac{r_a[\mathbf{r}]_\times + [\mathbf{r} \times (\mathbf{I} - \mathbf{R})\mathbf{e}_i]_\times}{||\mathbf{r}||^2}\mathbf{R}, \qquad a = x, y, z$$

  - $\mathbf{I}$: identity matrix
  - $\mathbf{e}_i$: the i-th vector of the standard basis in $\mathbb{R}^3$

1. G. Gallego and A. Yezzi, “A Compact Formula for the Derivative of a 3-D Rotation in Exponential Coordinates,” JMIV, vol. 51, no. 3, pp. 378–384, 2015.

Synthetic Dataset

Templates[1]

Background Images[2]

Test Images

1. Lieberknecht, Sebastian, et al. "A dataset and evaluation methodology for template-based tracking algorithms." ISMAR 2009
2. Jegou, Herve, Matthijs Douze, and Cordelia Schmid. "Hamming embedding and weak geometric consistency for large scale image search." ECCV, 2008.

Evaluated Algorithms

1. SIFT + OPnP
2. SIFT + IPPE
3. ASIFT + OPnP
4. ASIFT + IPPE
5. APE (Approximate Pose Estimation)
6. DPE (Direct Pose Estimation)

1. Lowe, David G. "Distinctive image features from scale-invariant keypoints." IJCV, 2004.
2. Morel, Jean-Michel, and Guoshen Yu. "ASIFT: A new framework for fully affine invariant image comparison." SIIMS, 2009.
3. Zheng, Yinqiang, et al. "Revisiting the pnp problem: A fast, general and optimal solution." ICCV, 2013.
4. Collins, Toby, and Adrien Bartoli. "Infinitesimal plane-based pose estimation." IJCV, 2014

Evaluated Metric

$\tilde{\mathbf{p}} \equiv (\tilde{\mathbf{R}}, \tilde{\mathbf{t}})$: ground-truth pose
$\mathbf{p} \equiv (\mathbf{R}, \mathbf{t})$: estimated pose

– Rotation error (degree)

$$E_r = \mathrm{acosd}\left(\frac{\mathrm{Tr}\left(\mathbf{R}^{\mathrm{T}} \cdot \tilde{\mathbf{R}}\right)}{2}\right)$$

– Translation error (%)

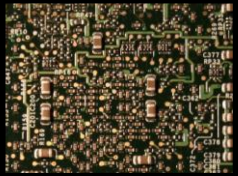
$$E_t = \frac{||\tilde{\mathbf{t}} - \mathbf{t}||}{||\tilde{\mathbf{t}}||} \times 100$$

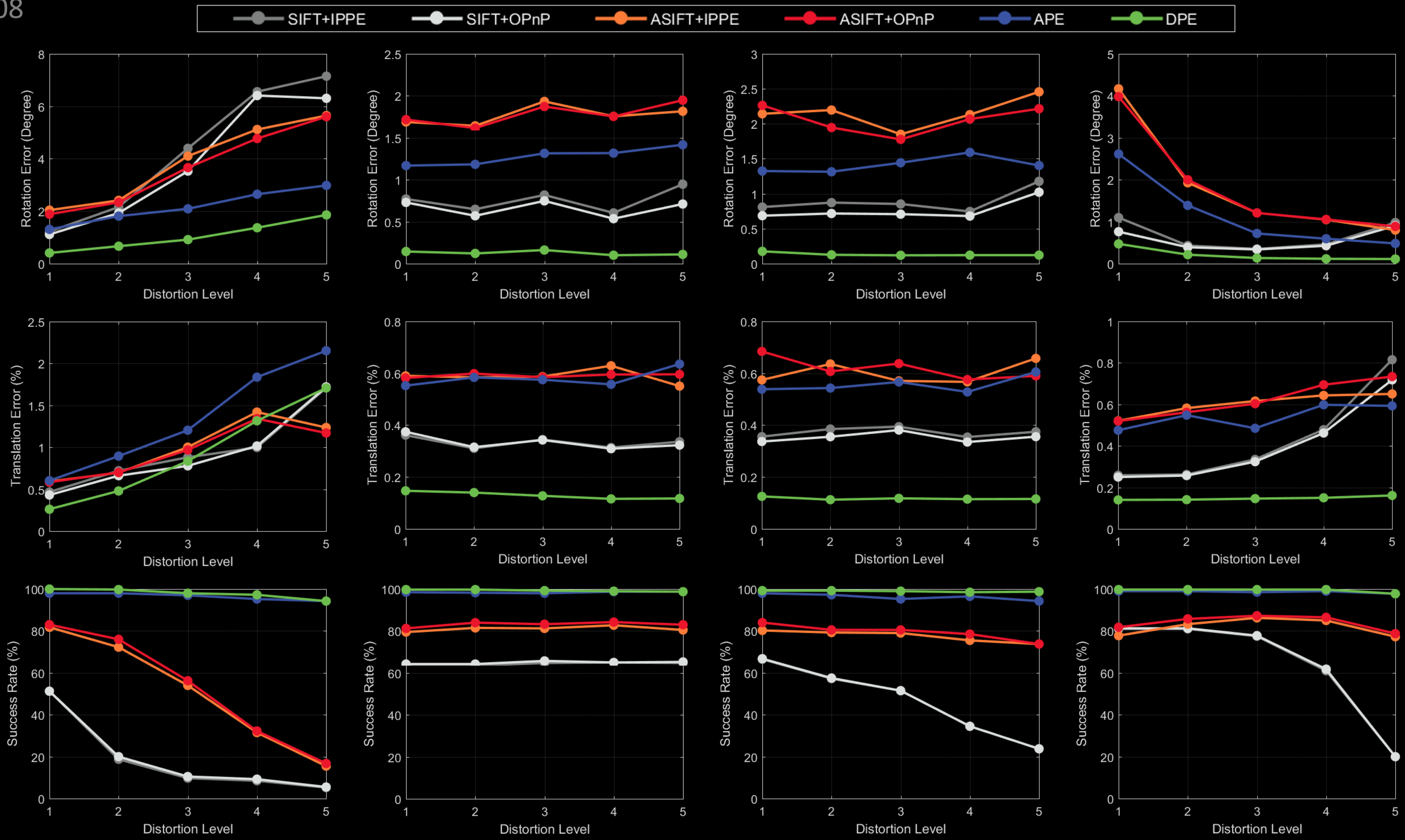
– Success rate (%)

- The percentage of poses that $E_r < 20^{\circ}$ and $E_t < 10\%$

Evaluation Results

- Results with undistorted test images.

																								
Method	E_r	E_t	SR	E_r	E_t	SR	E_r	E_t	SR	E_r	E_t	SR	E_r	E_t	SR	E_r	E_t	SR	E_r	E_t	SR	E_r	E_t	SR
SIFT+IPPE	0.85	0.34	40.0	1.90	0.54	96.0	0.23	0.25	28.0	0.32	0.24	86.0	0.74	0.35	92.0	0.56	0.40	98.0	1.15	0.50	30.0	0.28	0.37	96.0
SIFT+OPnP	0.76	0.40	40.0	1.18	0.46	96.0	0.20	0.24	28.0	0.25	0.24	86.0	0.56	0.32	92.0	0.55	0.43	98.0	1.48	0.47	30.0	0.25	0.36	96.0
ASIFT+IPPE	9.70	2.92	20.0	2.96	0.81	94.0	1.48	0.43	**100**	1.65	0.51	94.0	1.59	0.57	**100**	1.29	0.34	98.0	2.17	0.52	52.0	1.96	0.36	90.0
ASIFT+OPnP	8.20	2.22	22.0	2.72	0.74	**100**	1.38	0.41	**100**	1.53	0.45	96.0	1.40	0.50	98.0	1.26	0.35	**100**	1.33	0.37	52.0	1.80	0.36	94.0
APE	1.10	0.33	**100**	1.44	0.42	**100**	0.90	0.47	98.0	2.56	1.23	94.0	1.03	0.35	**100**	1.63	0.49	**100**	1.96	0.91	**100**	1.57	0.68	98.0
DPE	**0.39**	**0.17**	**100**	**0.42**	**0.24**	**100**	**0.16**	**0.14**	**100**	**0.16**	**0.12**	**98.0**	**0.21**	**0.16**	**100**	**0.21**	**0.11**	**100**	**0.15**	**0.14**	**100**	**0.17**	**0.13**	**100**



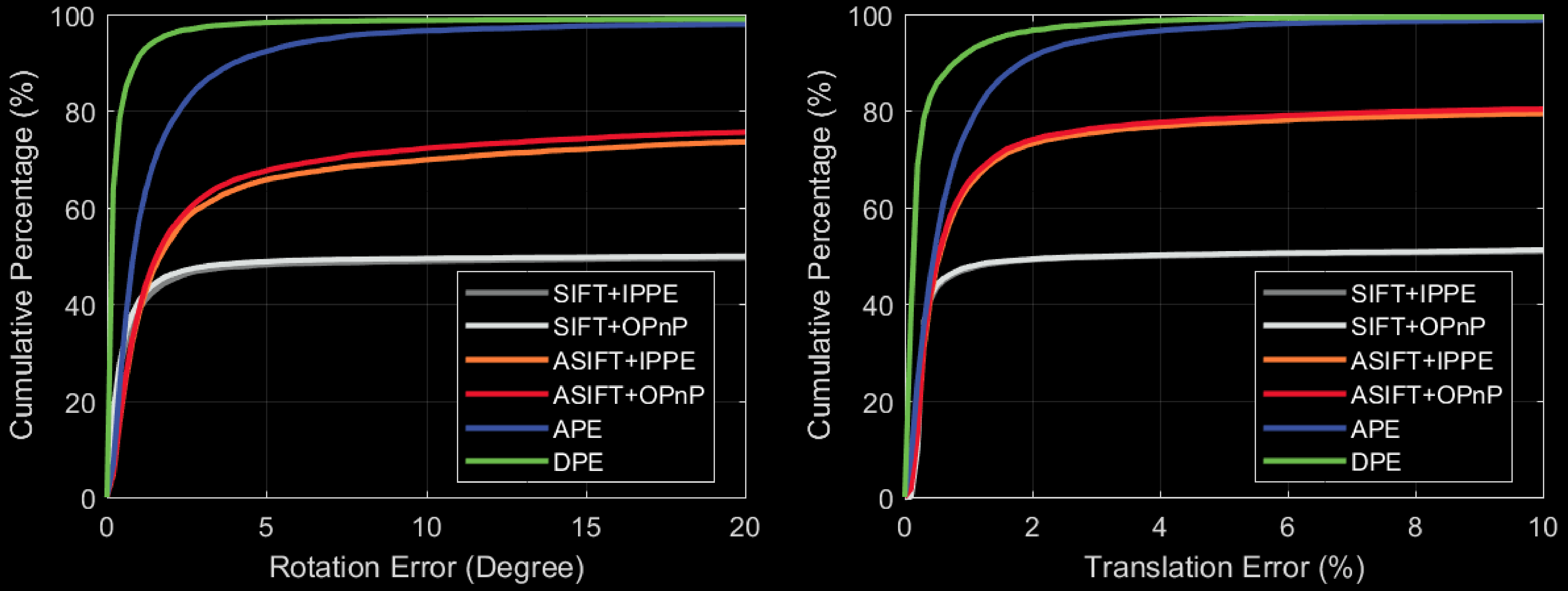
$E_r(°)$

$E_t(\%)$

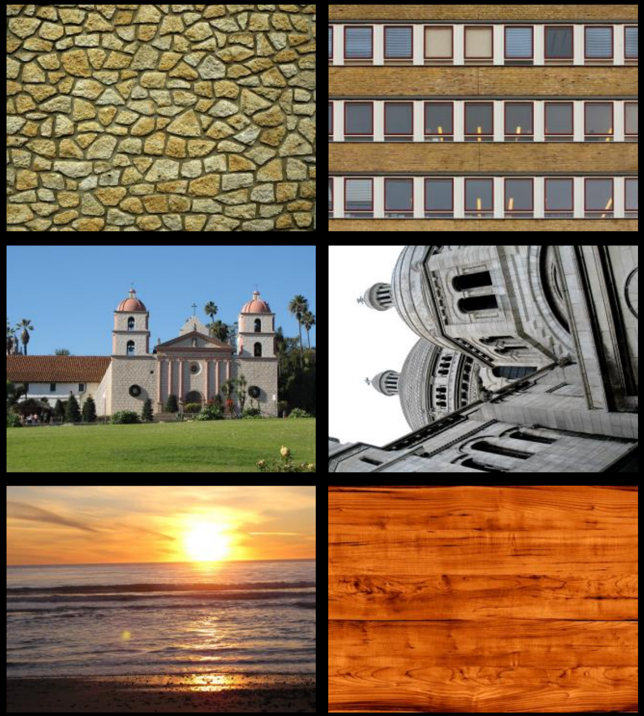
$\mathrm{SR}(\%)$

(a) Gaussian blur (b) JPEG compression (c) Intensity change (d) Tilt angle

Overall Evaluation



Visual Tracking Dataset[1]

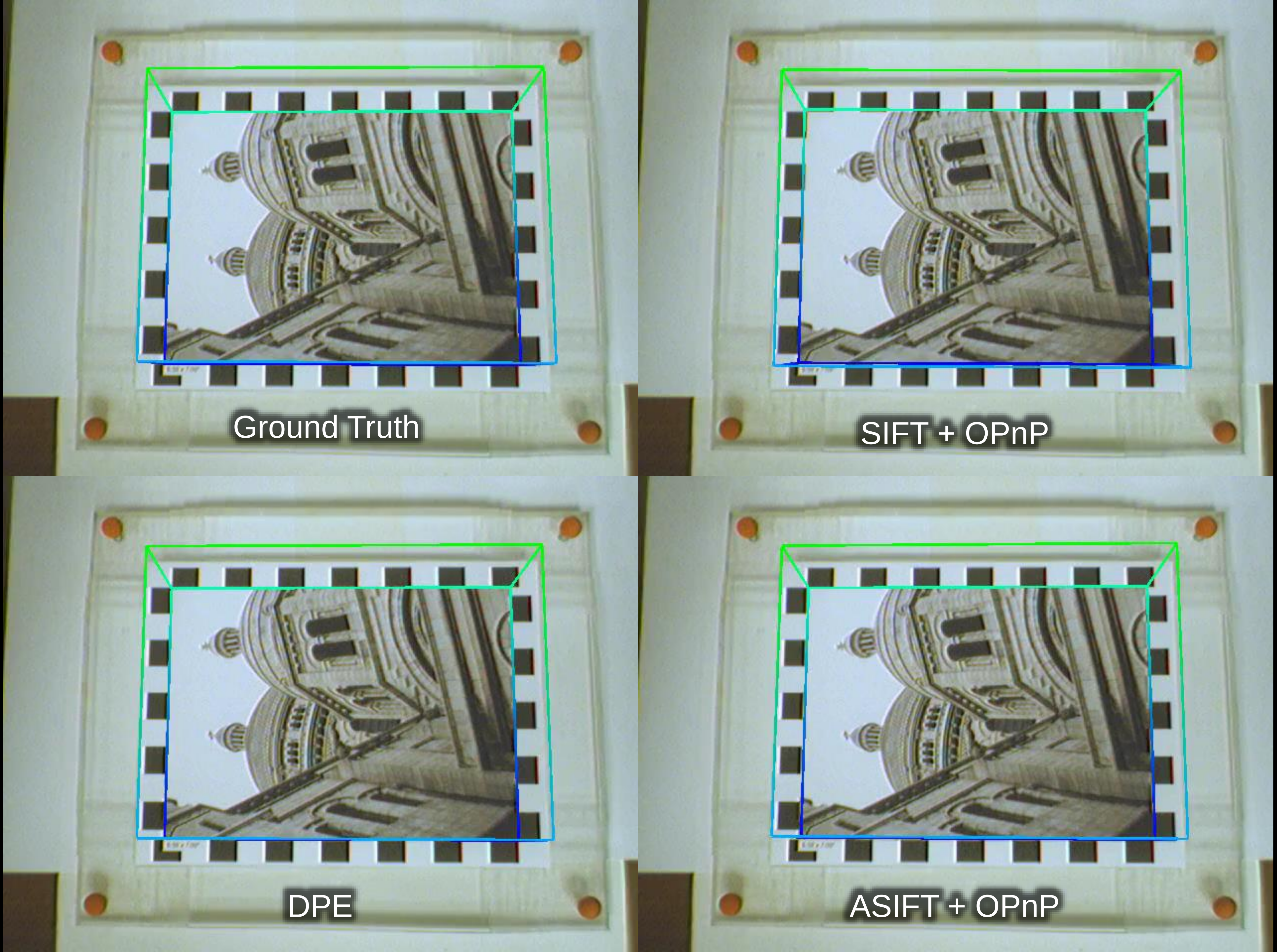


Templates

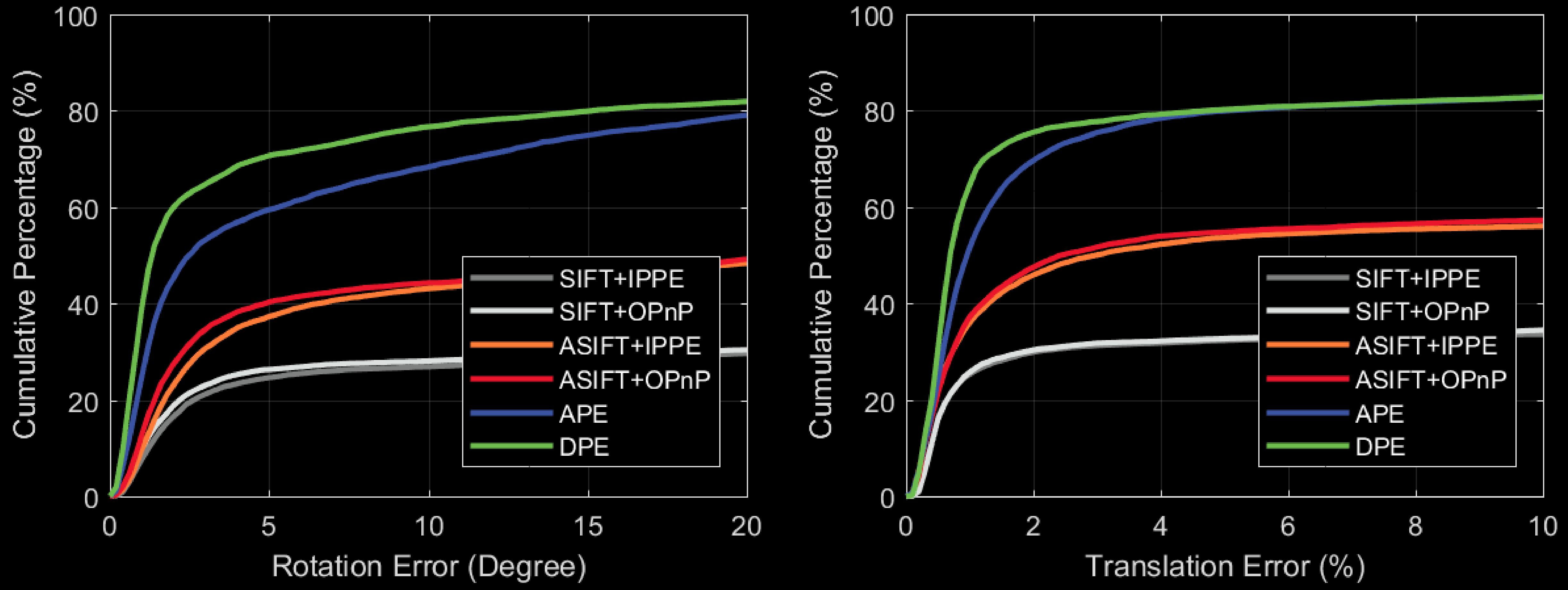
- Unconstrained
- Panning
- Rotation
- Perspective Distortion
- Zoom
- Static Lighting
- Dynamic Lighting
- Motion Blur x9

Motion Patterns

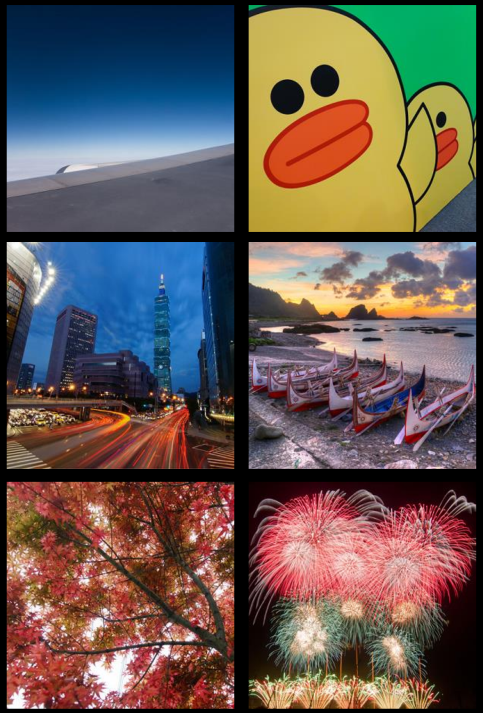
1. Gauglitz, Steffen, Tobias Höllerer, and Matthew Turk. "Evaluation of interest point detectors and feature descriptors for visual tracking." IJCV, 2011


Ground Truth
SIFT + OPnP
DPE
ASIFT + OPnP

Overall Evaluation



OPT Dataset (Proposed)



Templates

×

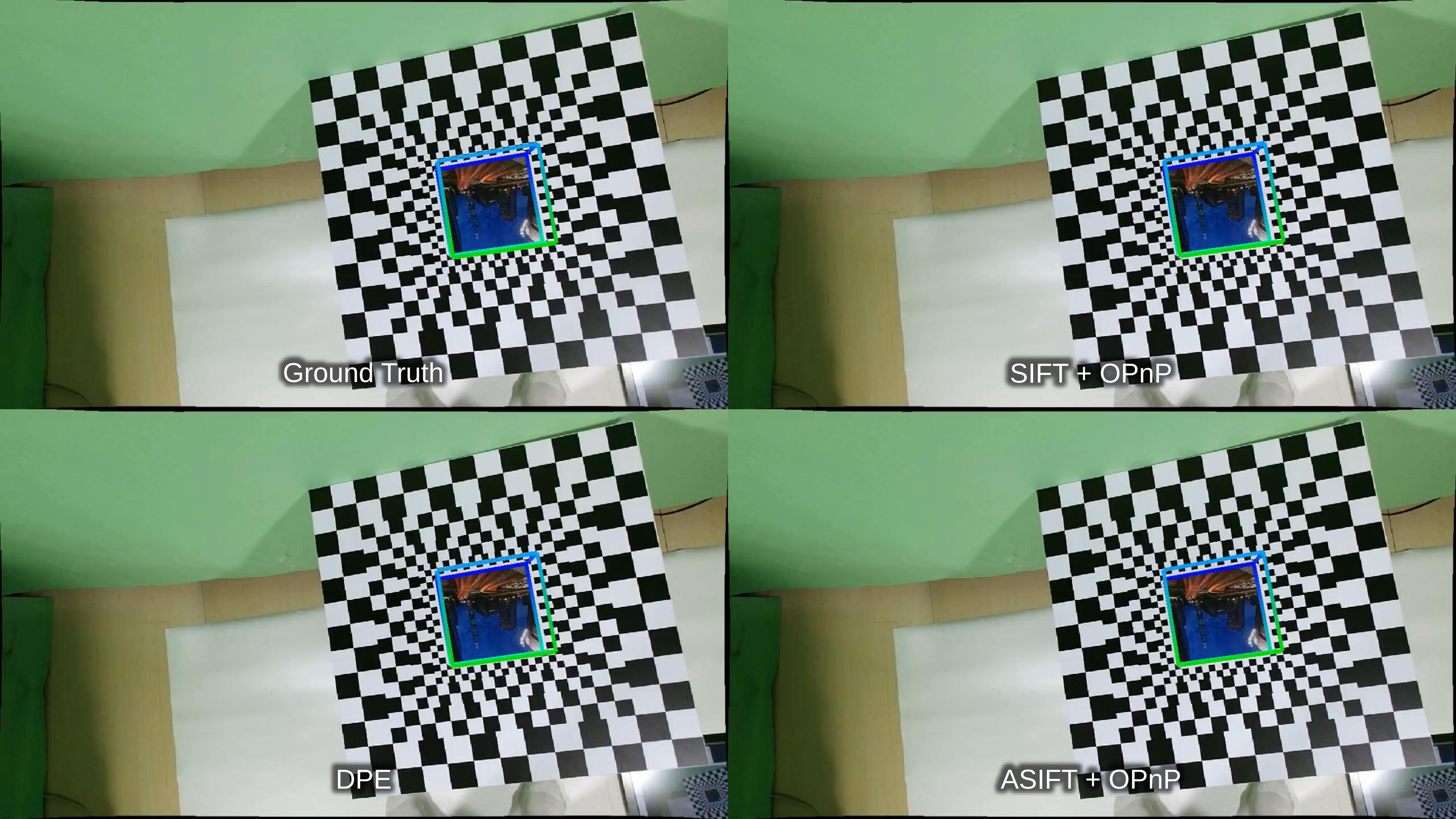
- Translation
- Zoom
- In-plane Rotation
- Out-of-plane Rotation
- Flashing Light
- Moving Light
- Free Motion

Motion Patterns

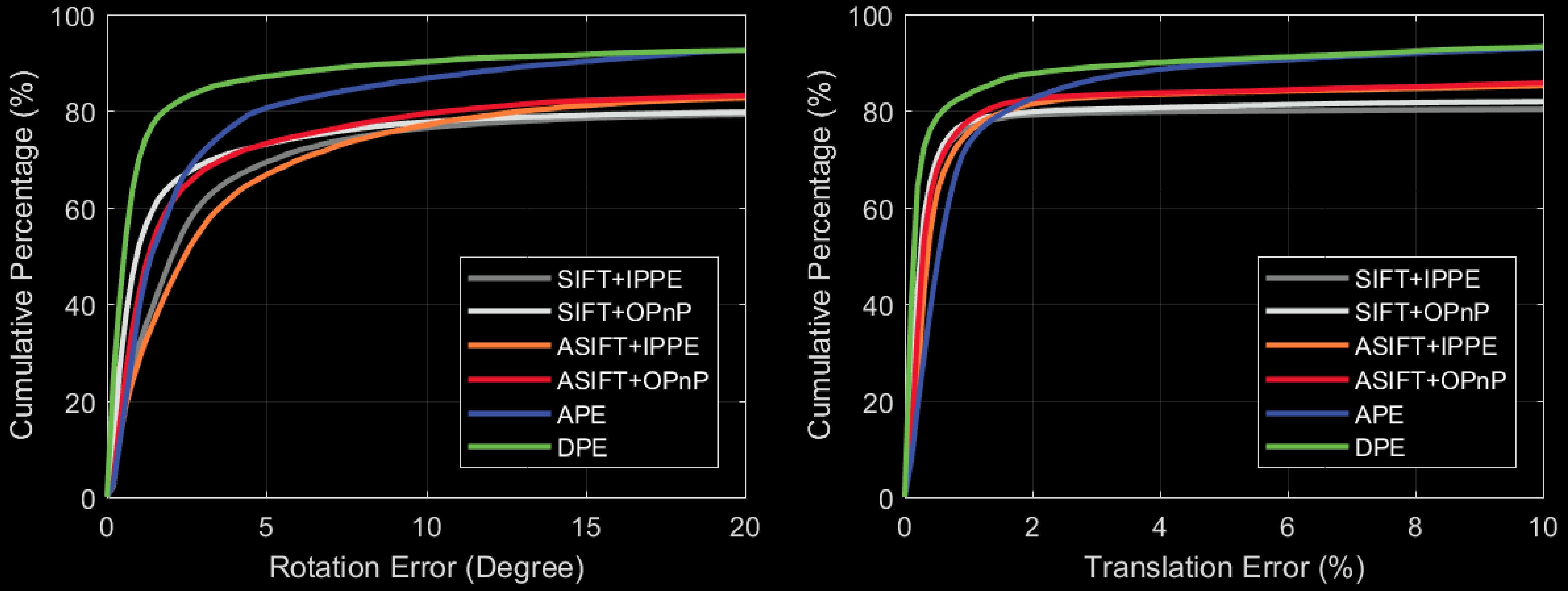
×

5

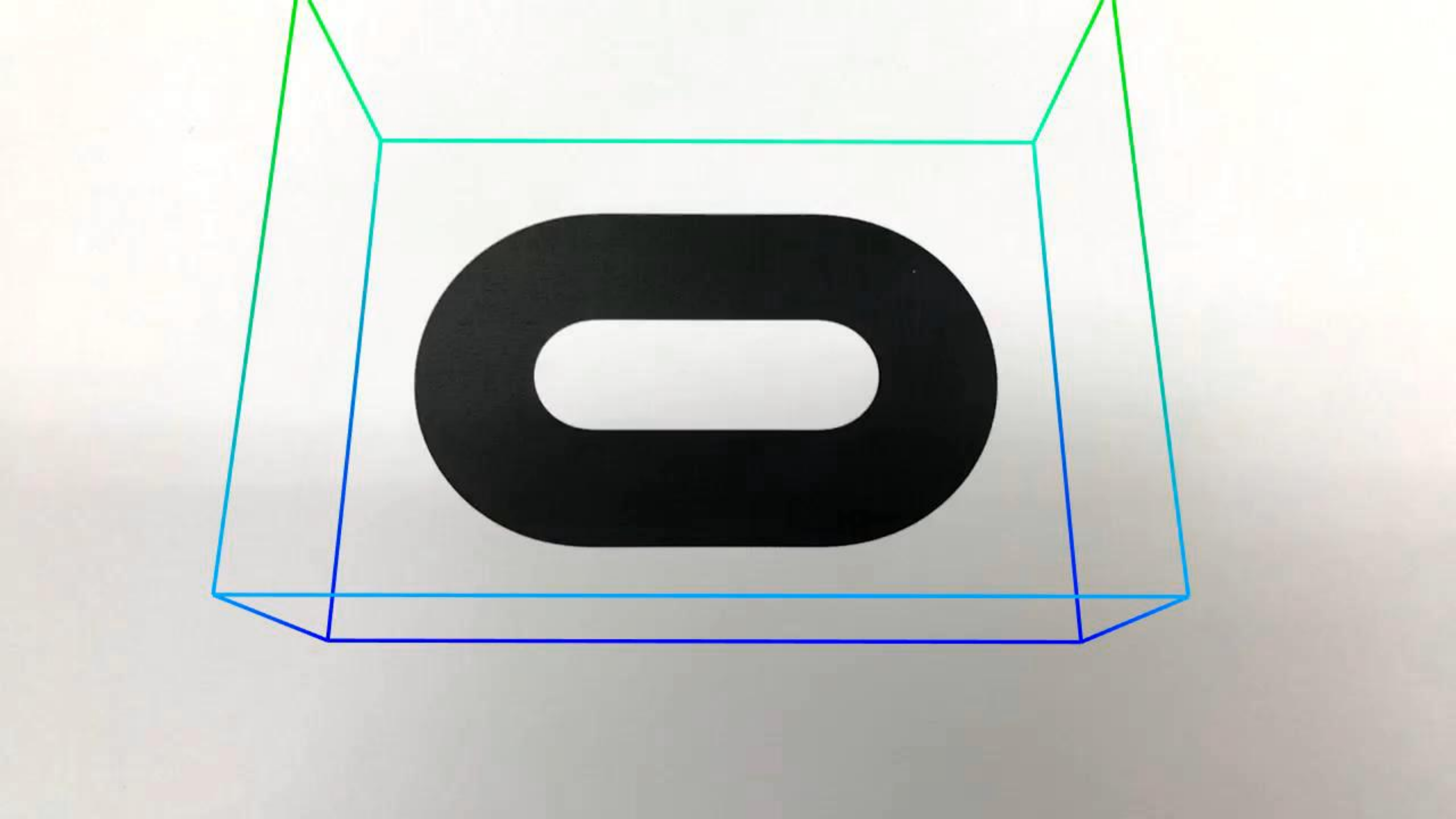
Speeds


Ground Truth
SIFT + OPnP
DPE
ASIFT + OPnP

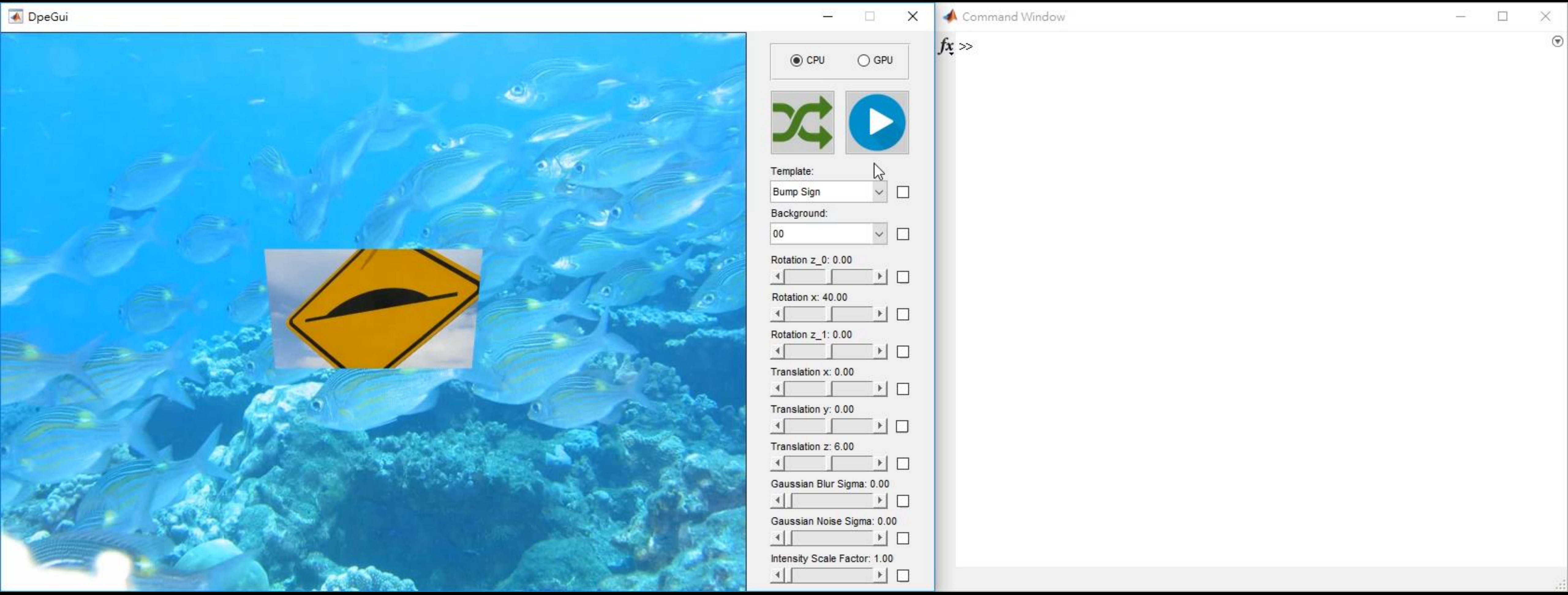
Overall Evaluation



How About This?



DPE Demo Video



Average Runtime

- Average runtime (measured in seconds) using MATLAB.
  - Core i7-6700K 4.0 GHz processor
  - 32 GB RAM
  - NVIDIA GTX 970 GPU
- Numbers in parentheses denote the average runtime of the CUDA implementation.

Dataset	SIFT-based Approach	ASIFT-based Approach	DPE		
			APE	PR	Total
Synthetic	7.446	10.912	10.549 (**1.505**)	0.571 (**0.117**)	11.120 (**1.622**)
VT	3.618	15.814	17.920 (**1.217**)	0.694 (**0.180**)	18.615 (**1.397**)
OPT	11.364	38.944	18.545 (**0.994**)	0.214 (**0.088**)	18.759 (**1.082**)

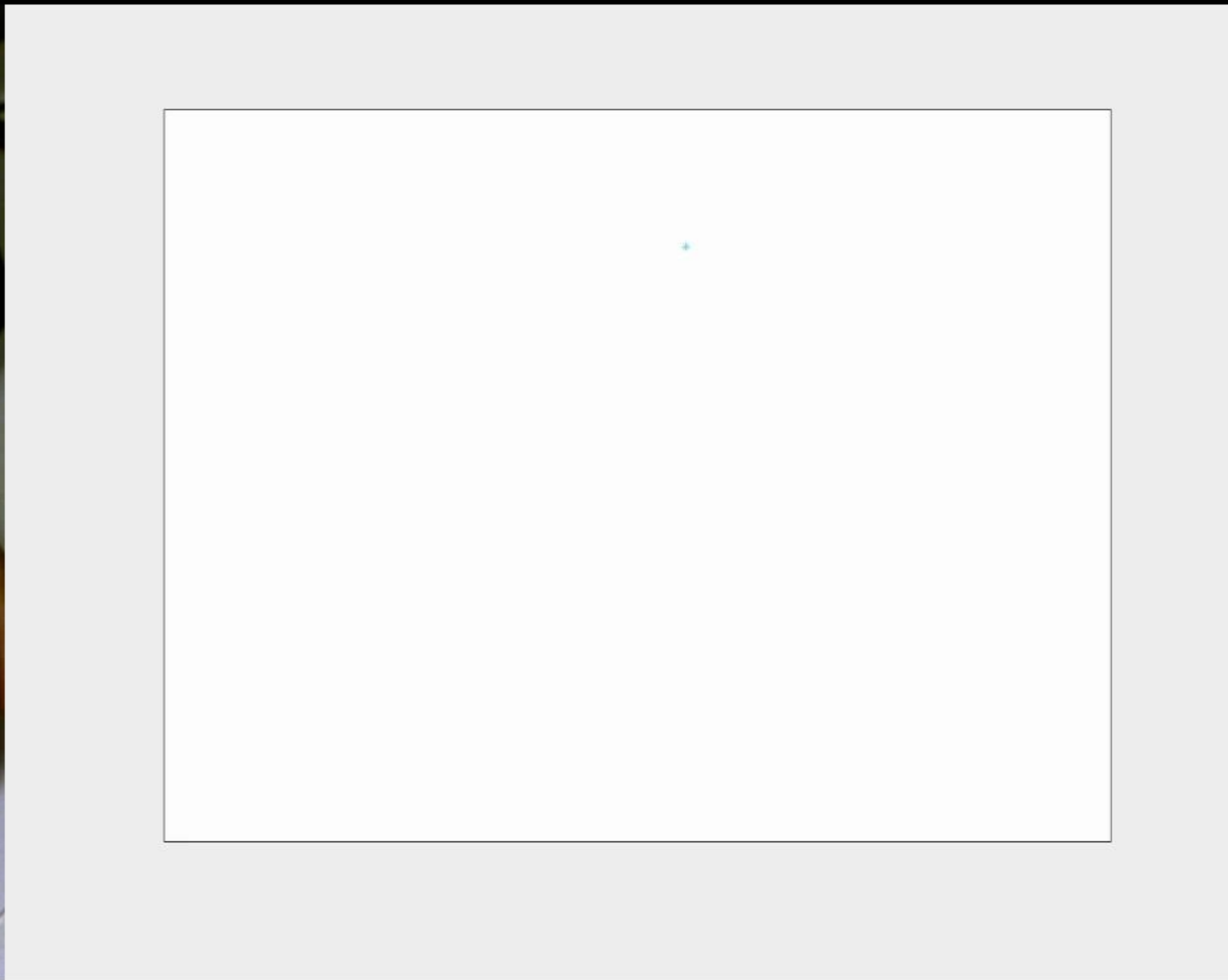
DodecaPen

Dodecahedron + Pen

6DoF Tracking

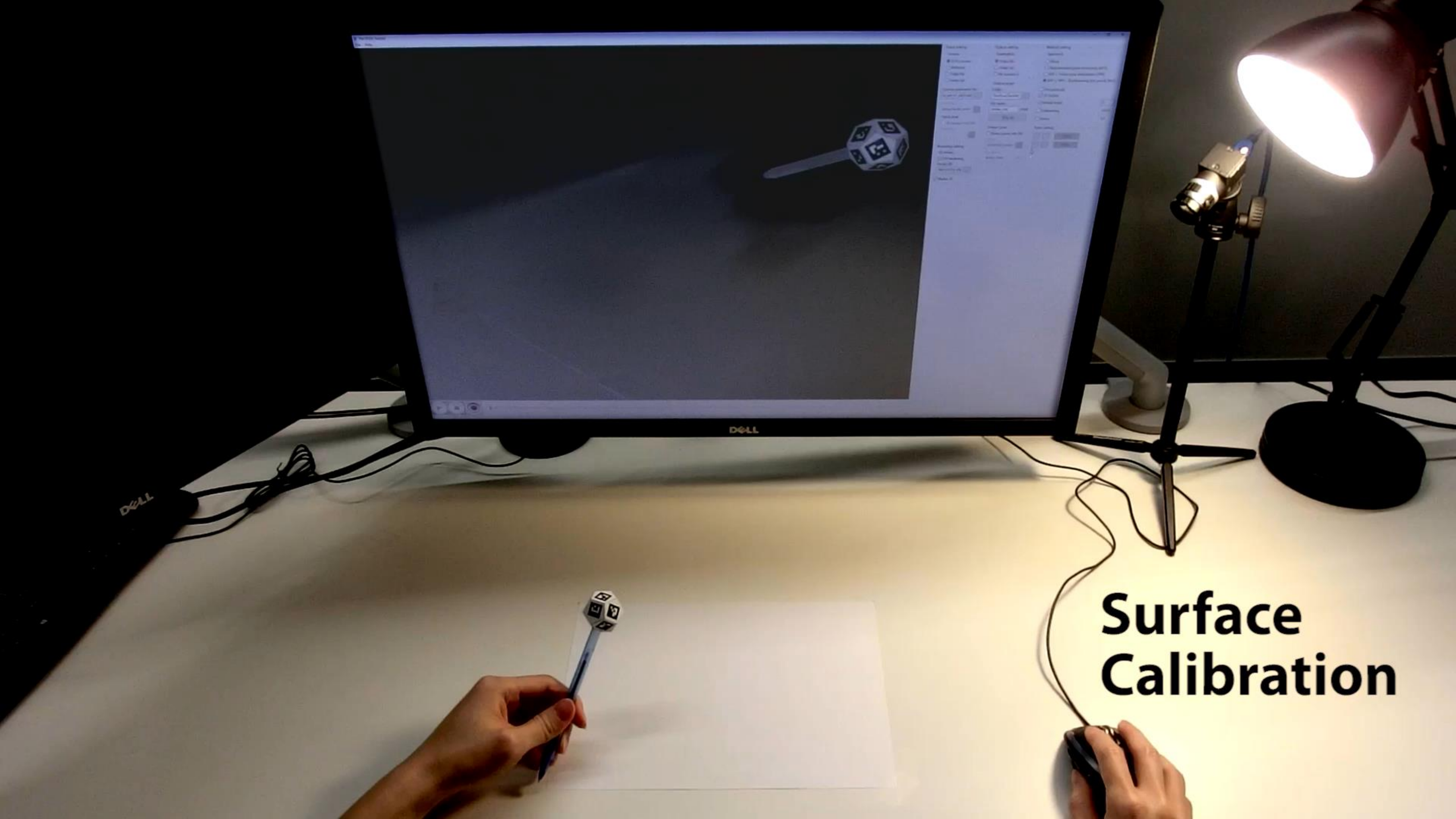
DodecaPen: Puppy

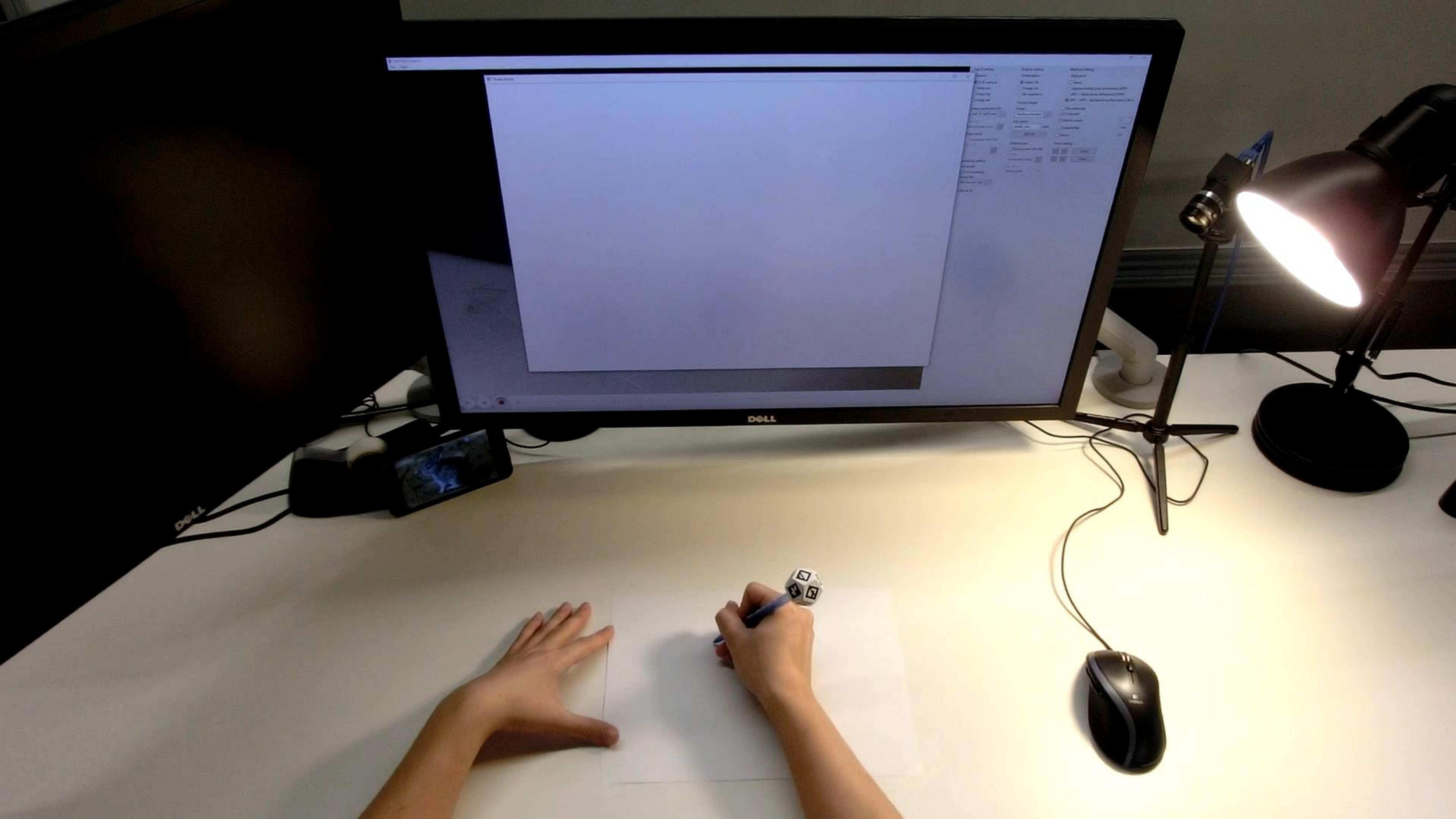
Input Frames



Pen-tip Trajectory

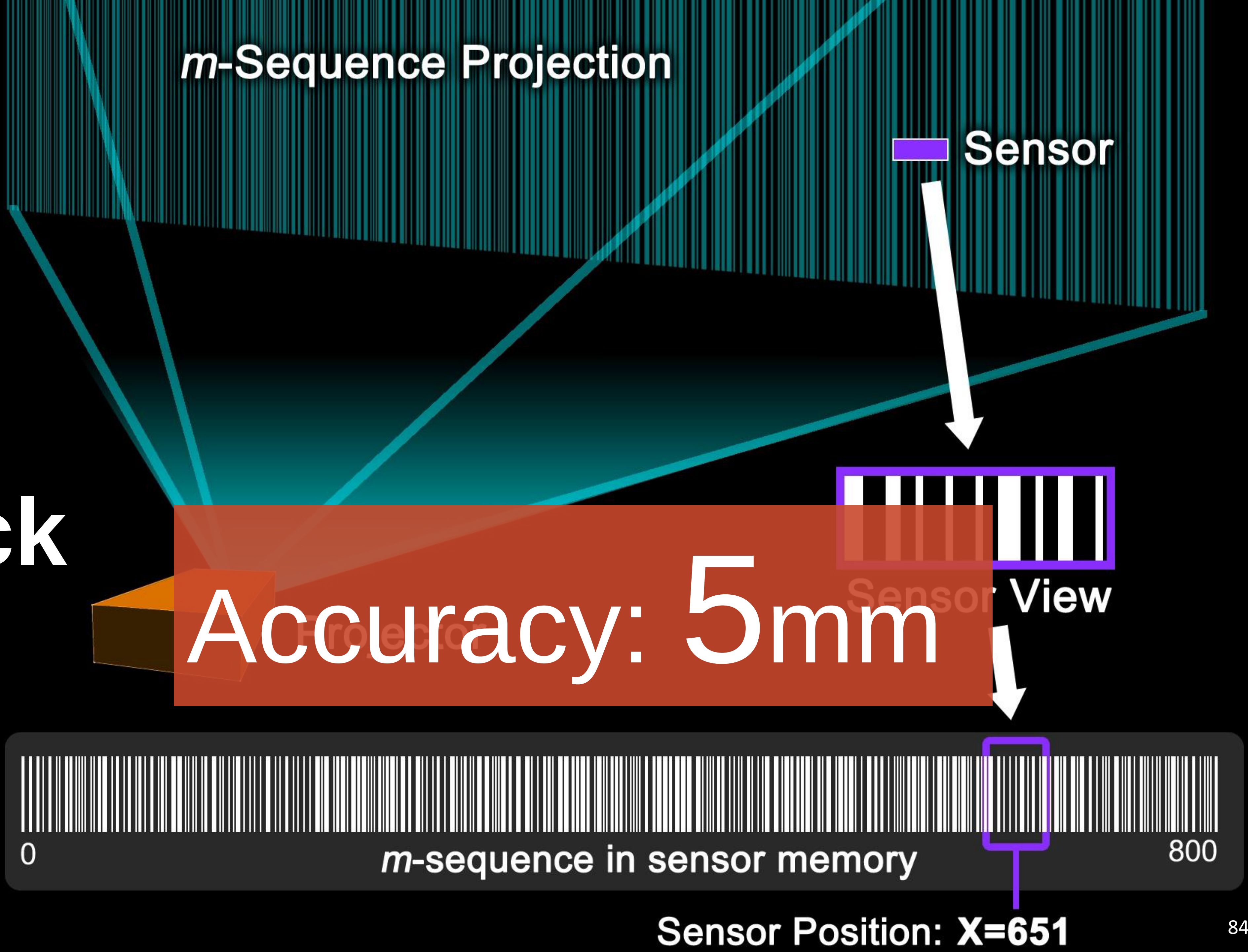
How To Use?


DELL
DELL
Surface
Calibration


DELL
DELL

Related Work

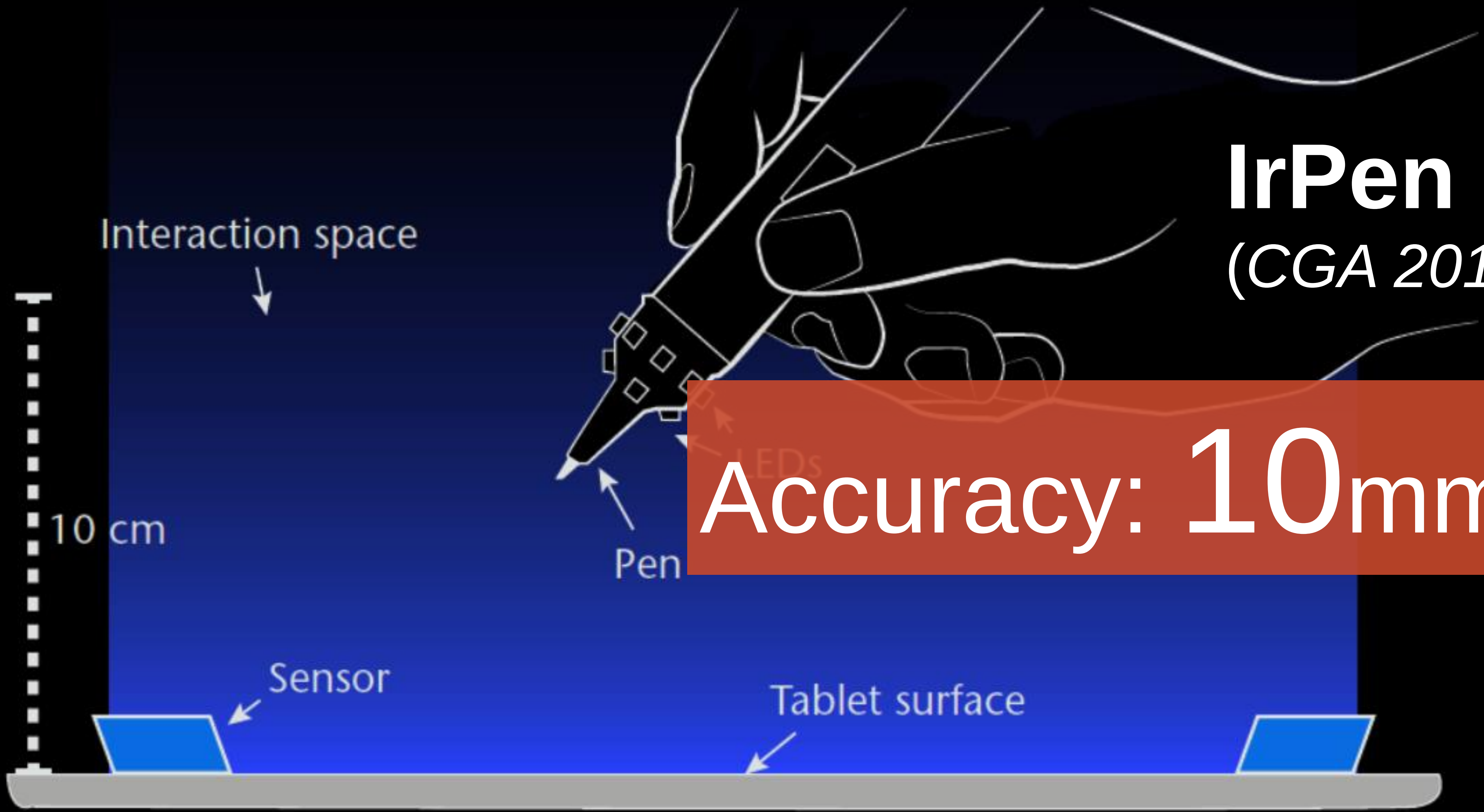
Lumitrack
(UIST 2013)



Accuracy: 5mm

IrPen

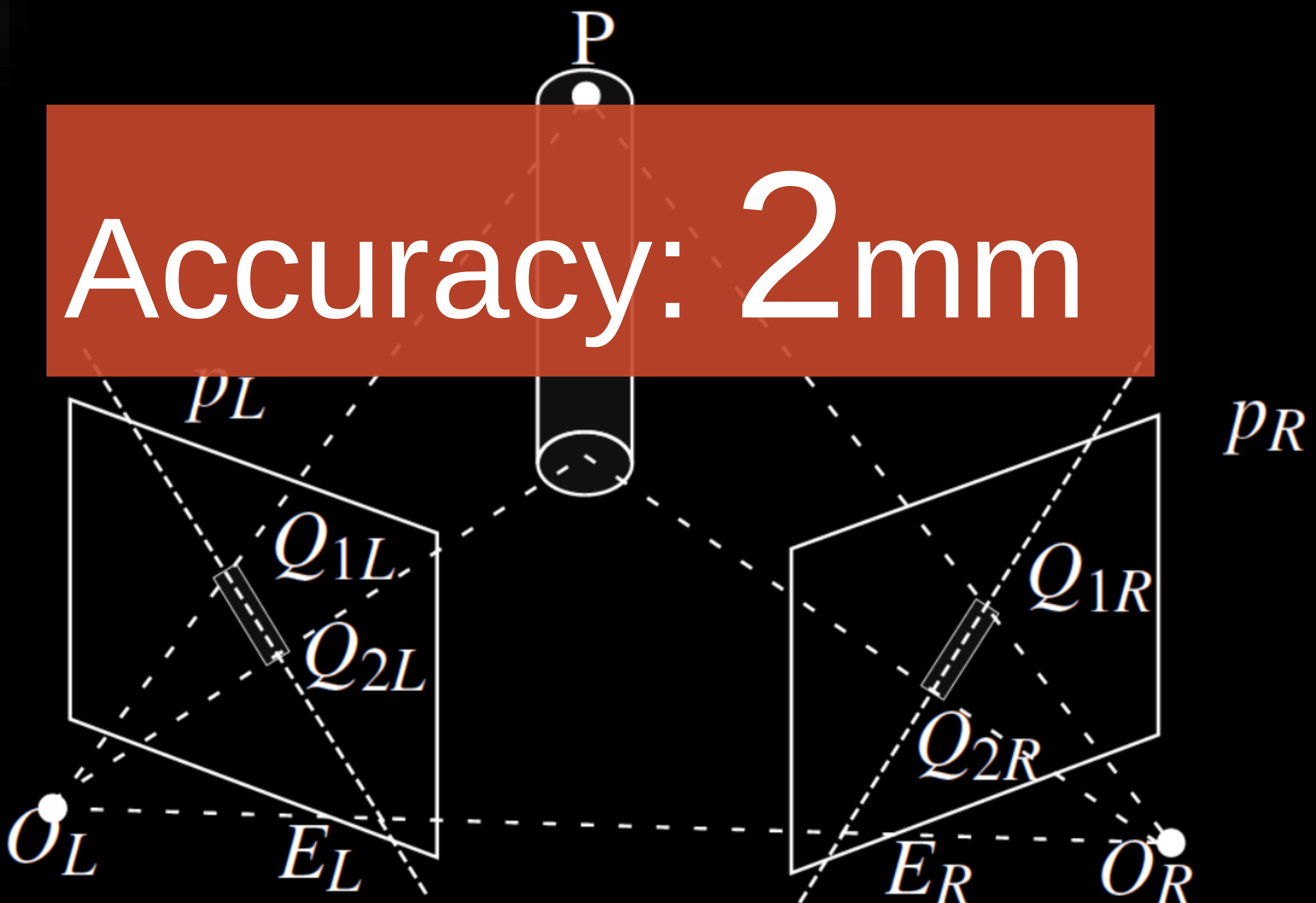
(CGA 2014)



Accuracy: 10mm

Light Chisel

(*PG 2015*)

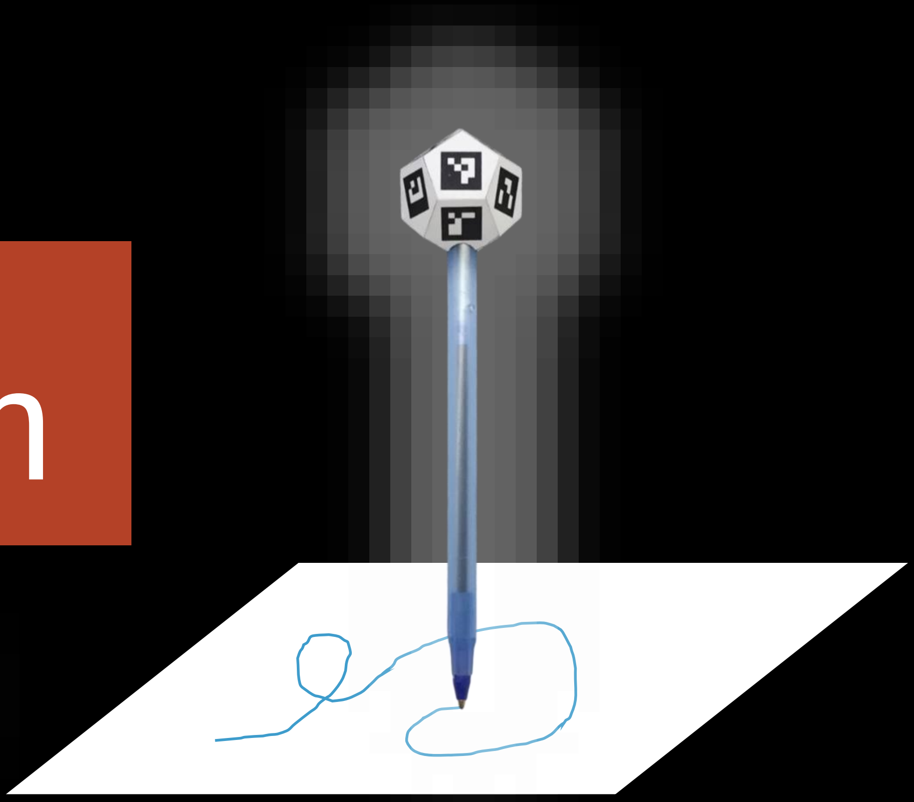


Accuracy: 2mm

DodecaPen

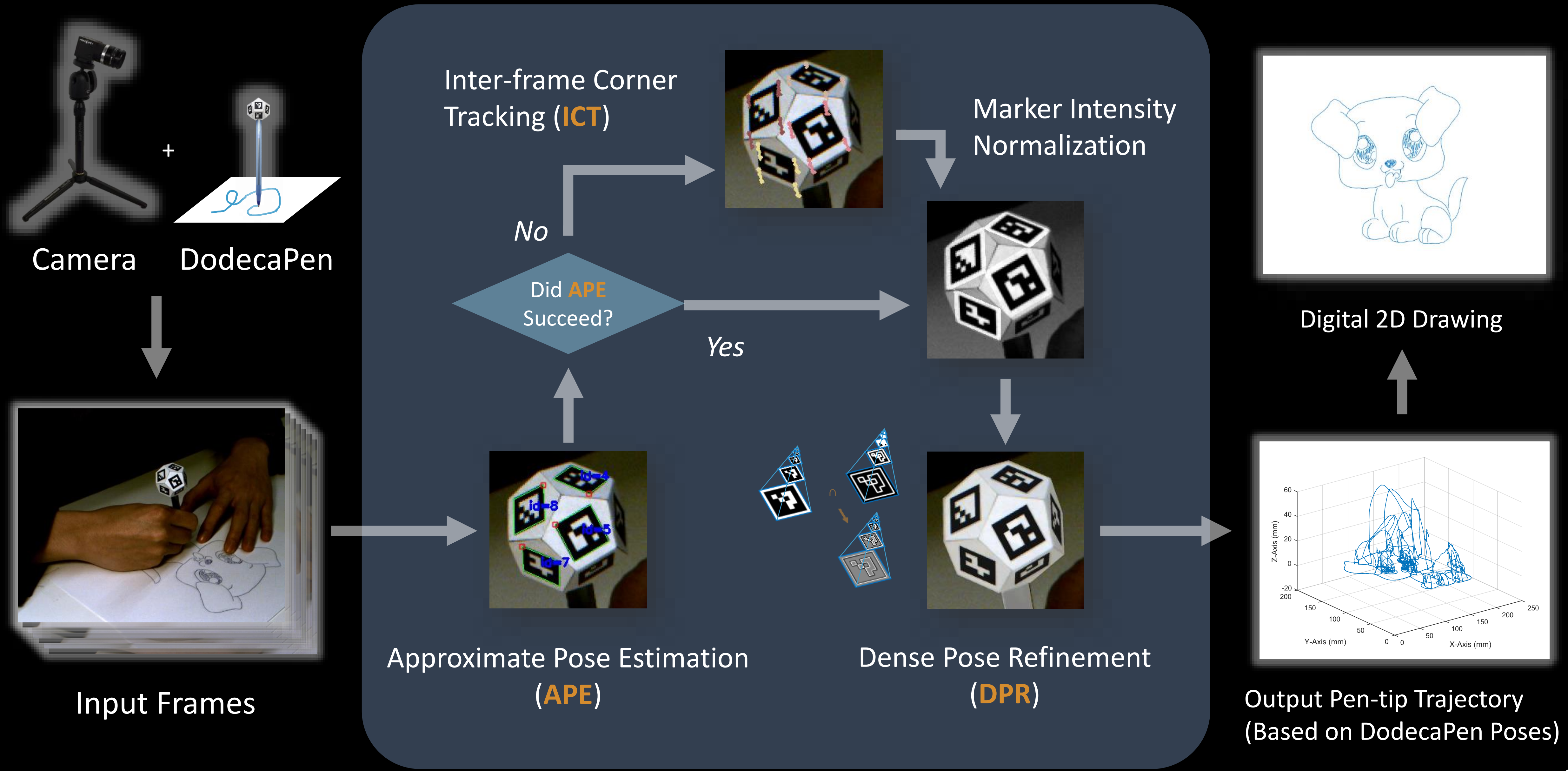
(Proposed)

Accuracy: 0.4mm

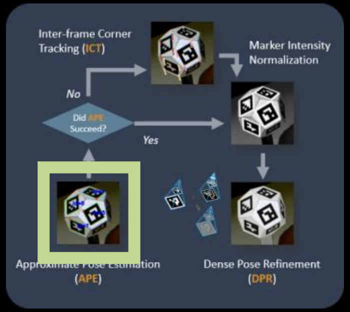


How To Implement?

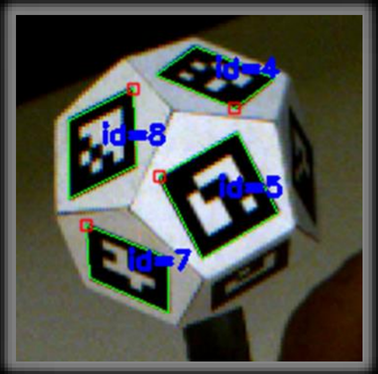
Proposed 6DoF Pose Tracking System



Approximate Pose Estimation (APE)



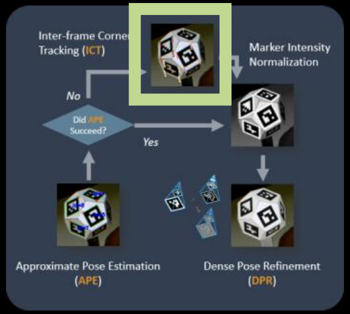
- Marker Detection
- Minimize reprojection error $E_r(\mathbf{p})$ with PnP algorithm to get the initial pose $\mathbf{p}'$



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

$\hat{\mathbf{u}}$: 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 $\mathbf{R}$ and translation vector $\mathbf{t}$)

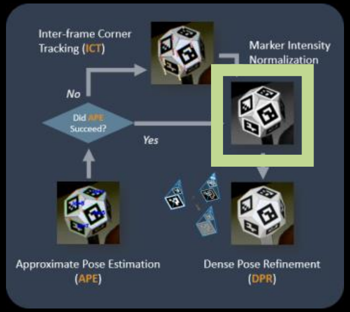
Inter-frame Corner Tracking (ICT)



If APE does not succeed…

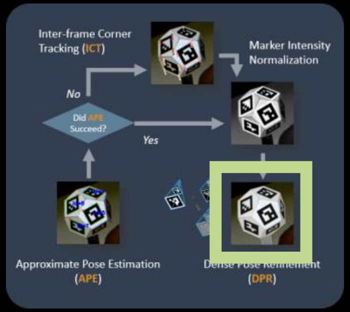
- Pyramidal Lucas-Kanade marker corner tracking
- PnP algorithm to get the initial pose $\mathbf{p}'$

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 $\mathbf{p}^*$

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

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 $\mathbf{t}$)

Gauss-Newton Iteration

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

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

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

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

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

1. Chain rule
2. Rotation vector $\mathbf{r}$
3. $\frac{\partial \mathbf{R}}{\partial \mathbf{r}}$

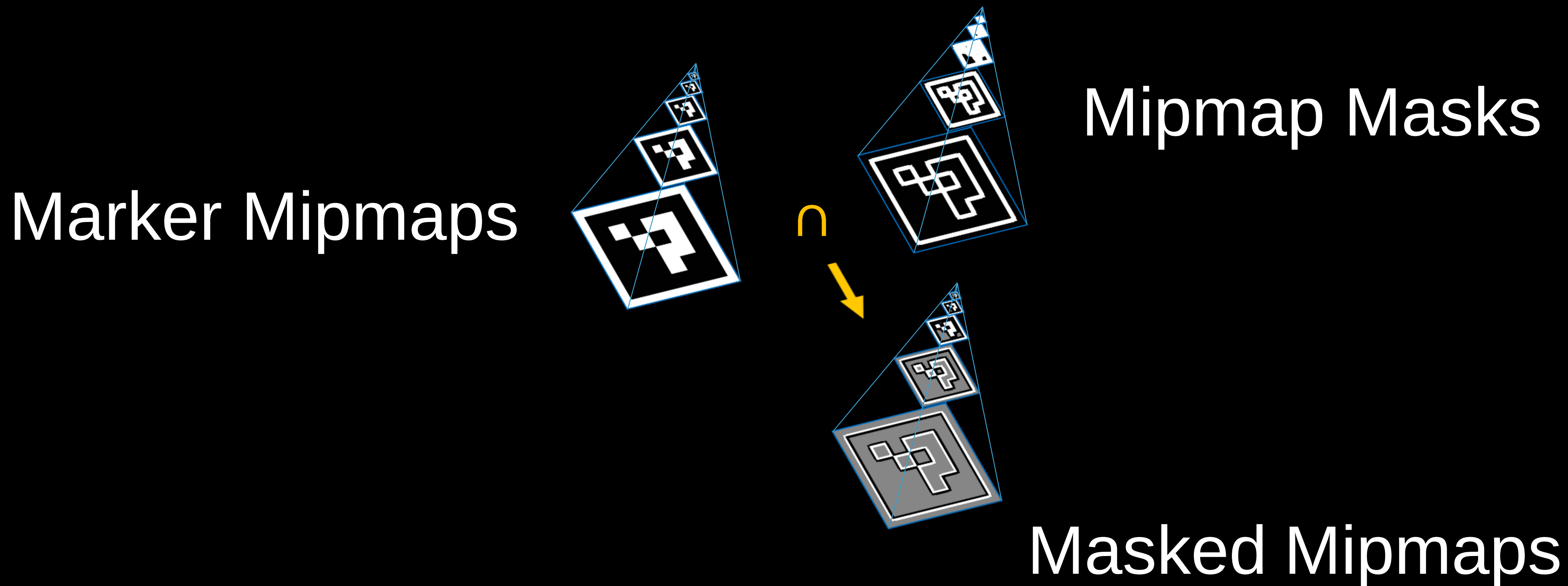
Backtracking Line Search (BLS)

- 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:

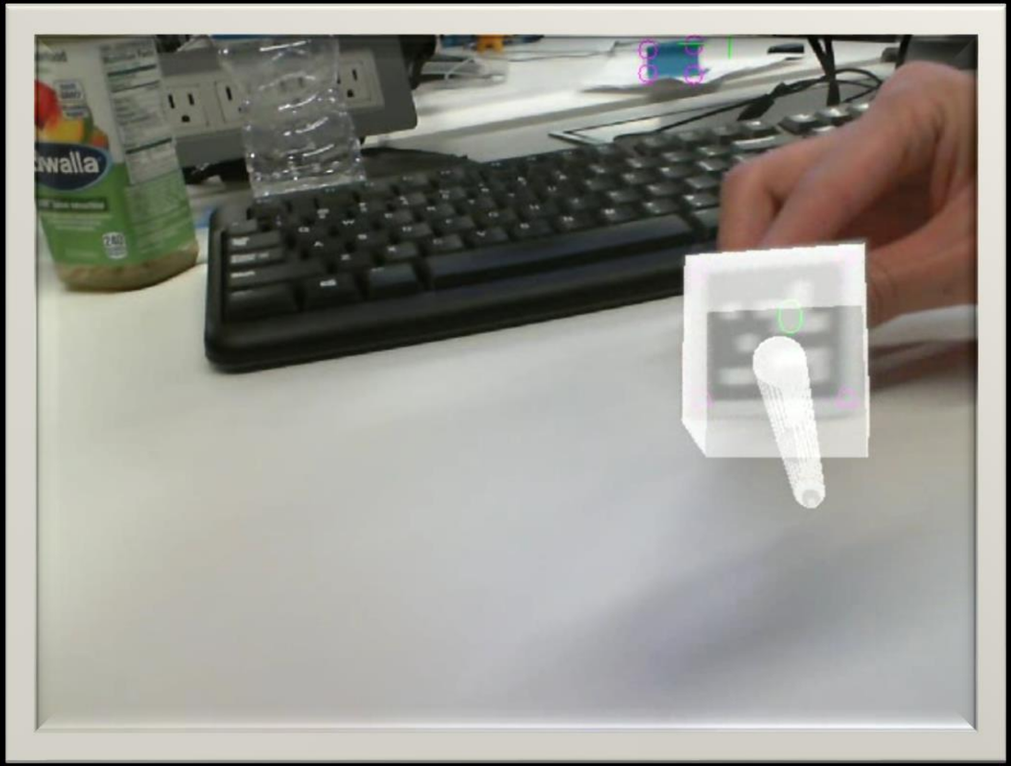
$$E_a(\mathbf{p}' + \Delta\mathbf{p}) \leq E_a(\mathbf{p}') + c\nabla E_a(\mathbf{p}')^{\mathrm{T}}\Delta\mathbf{p}$$

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

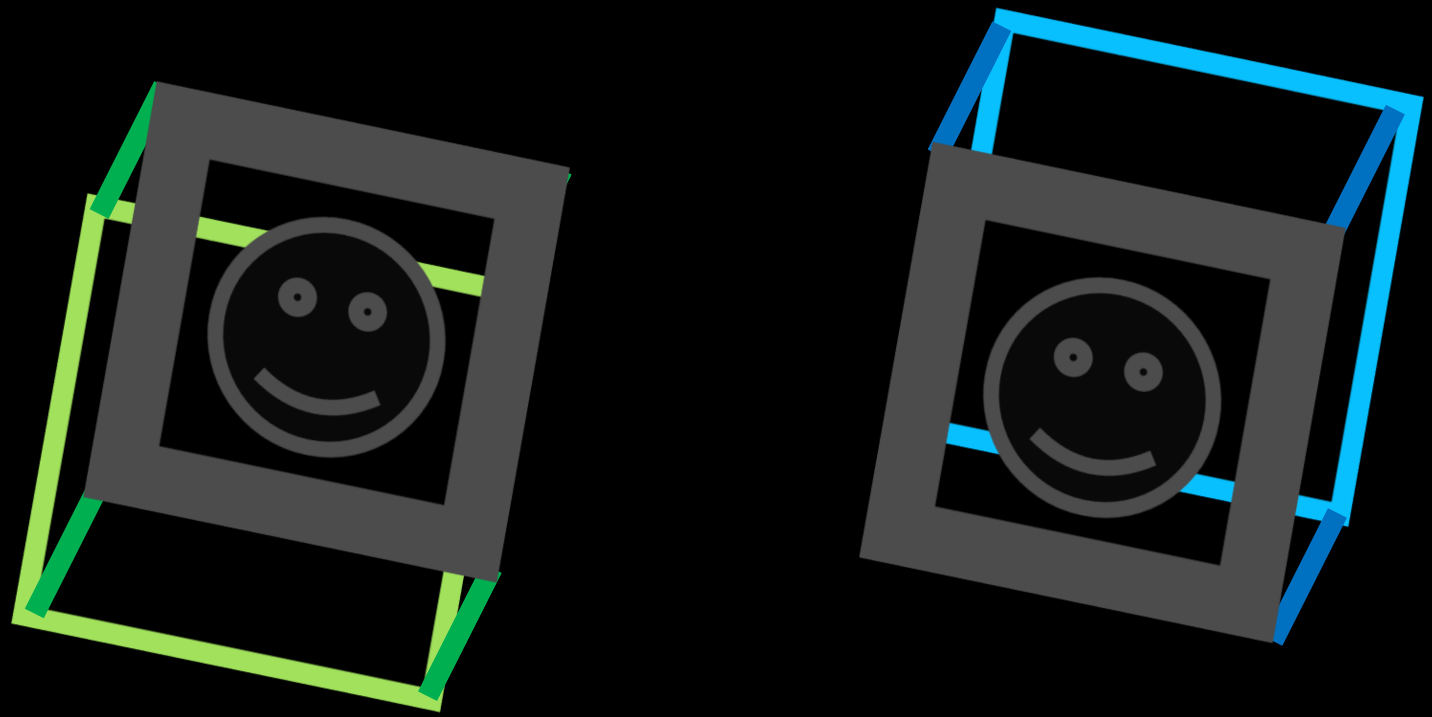
Masked Mipmaps



Why Dodecahedron?



Pose Jumping!

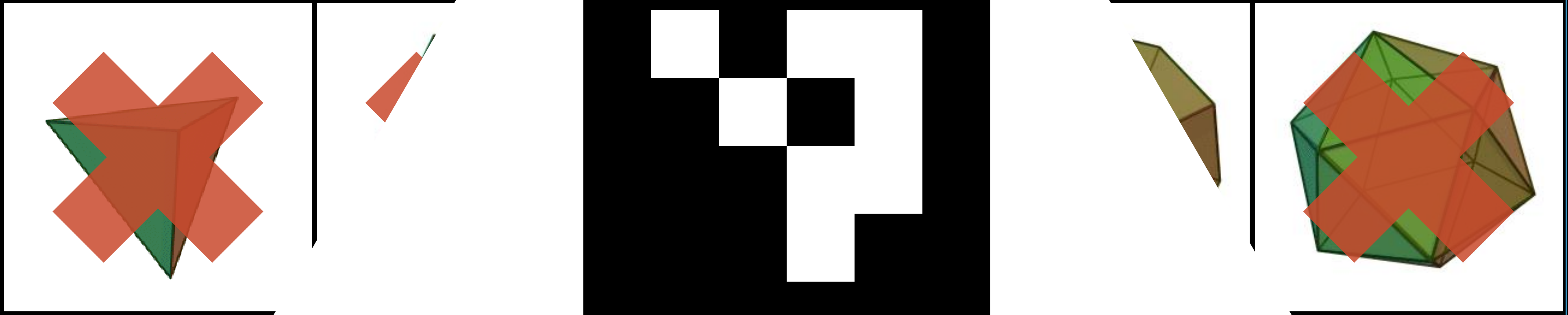


Multiple Candidates due to Coplanar Points

Platon c Solid

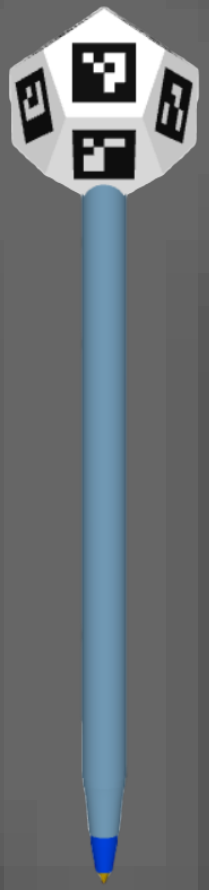
A Regular olyhedron

Tetrahedron	Cube	ecahedron	Icosahedron
Four faces	Six fa	e faces	Twenty faces



(source: wikipedia)

The Chosen One



Ideal ⇔ Real


Fail

Dodecahedron Calibration

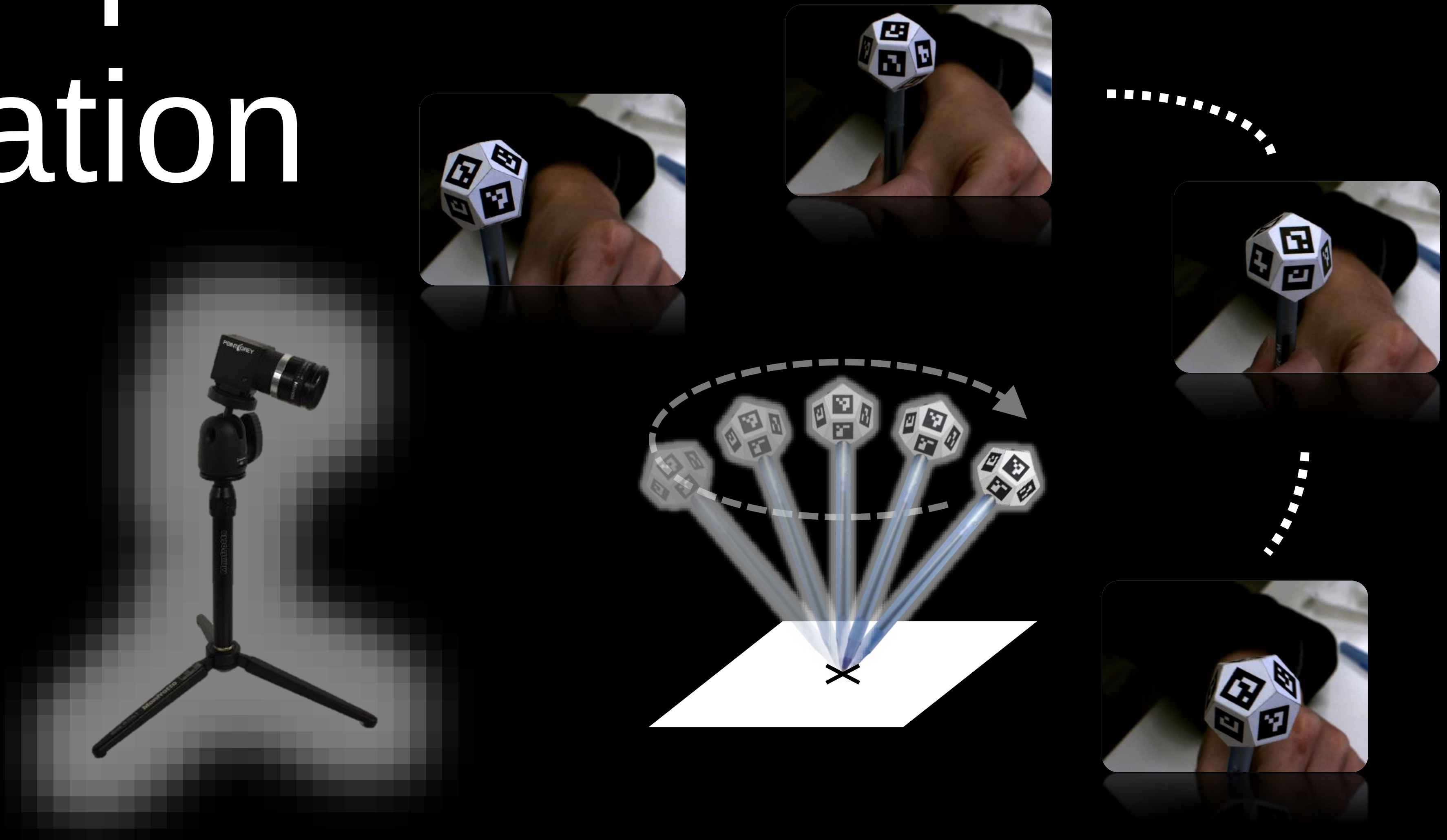
Dodecahedron Calibration (DC)

- 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

Pen-tip Calibration



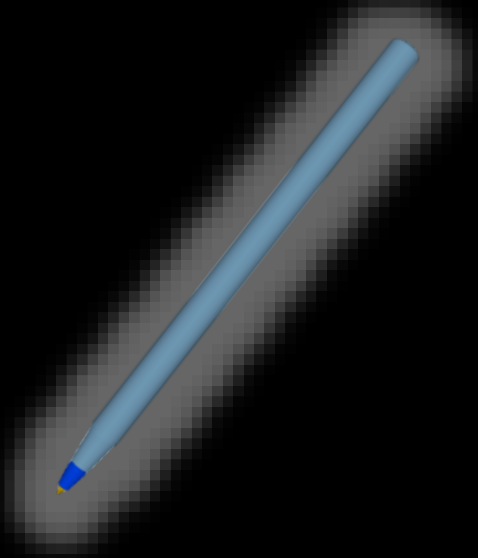
6 DoF Pose Tracking


DELL

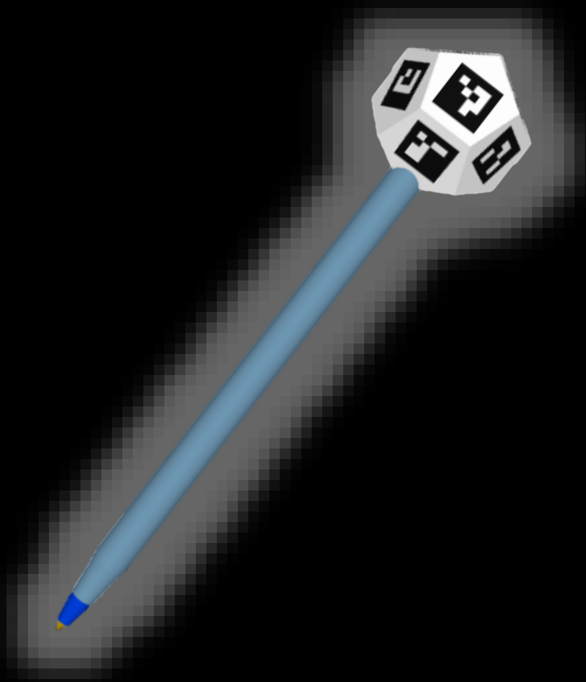

DELL
DELL


DELL
DELL

+



=



+

= ?


DELL
DELL
Logitech

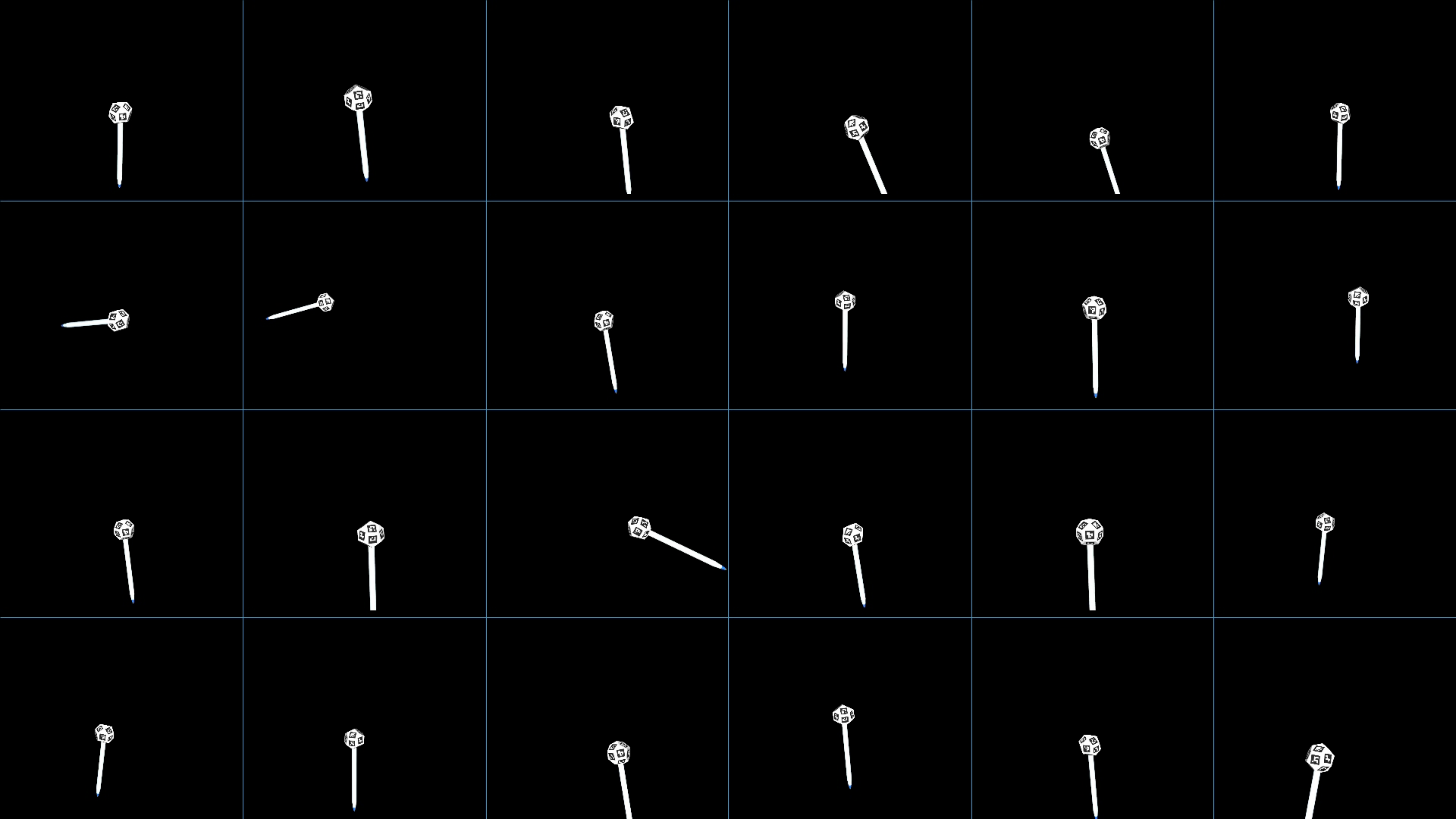

DELL
DELL

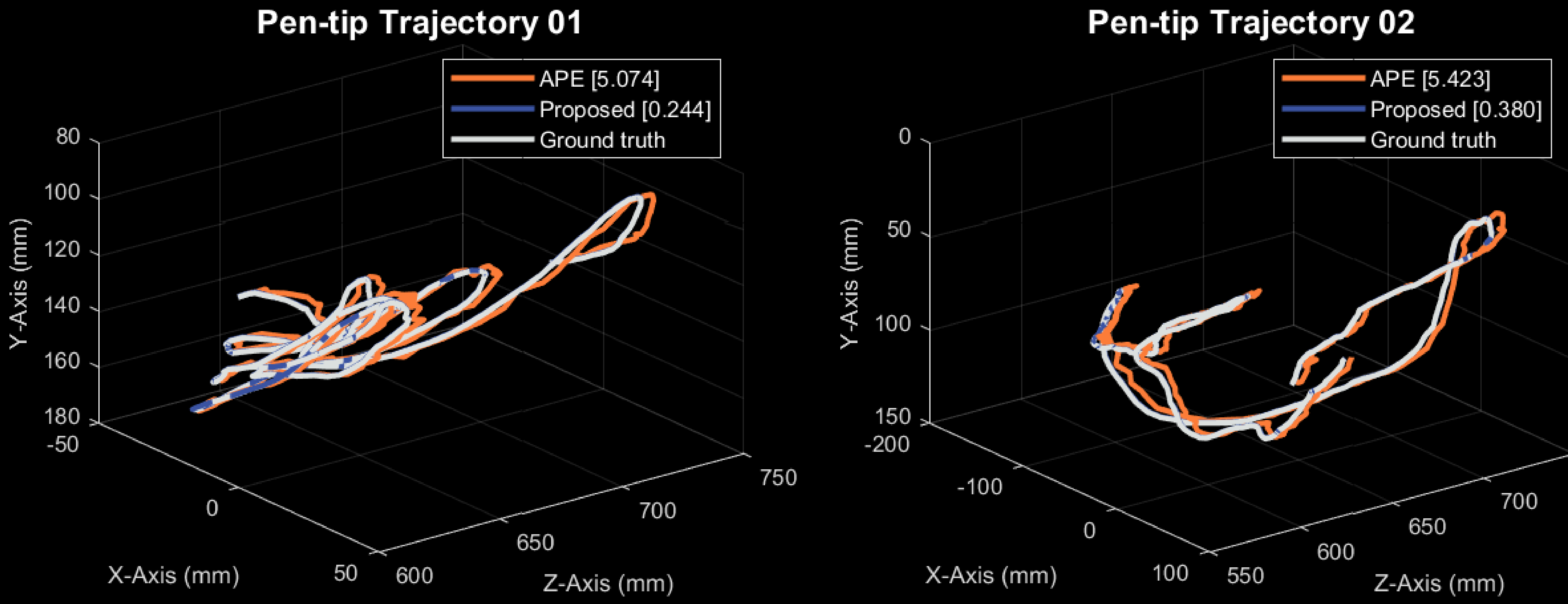
Performance Analysis

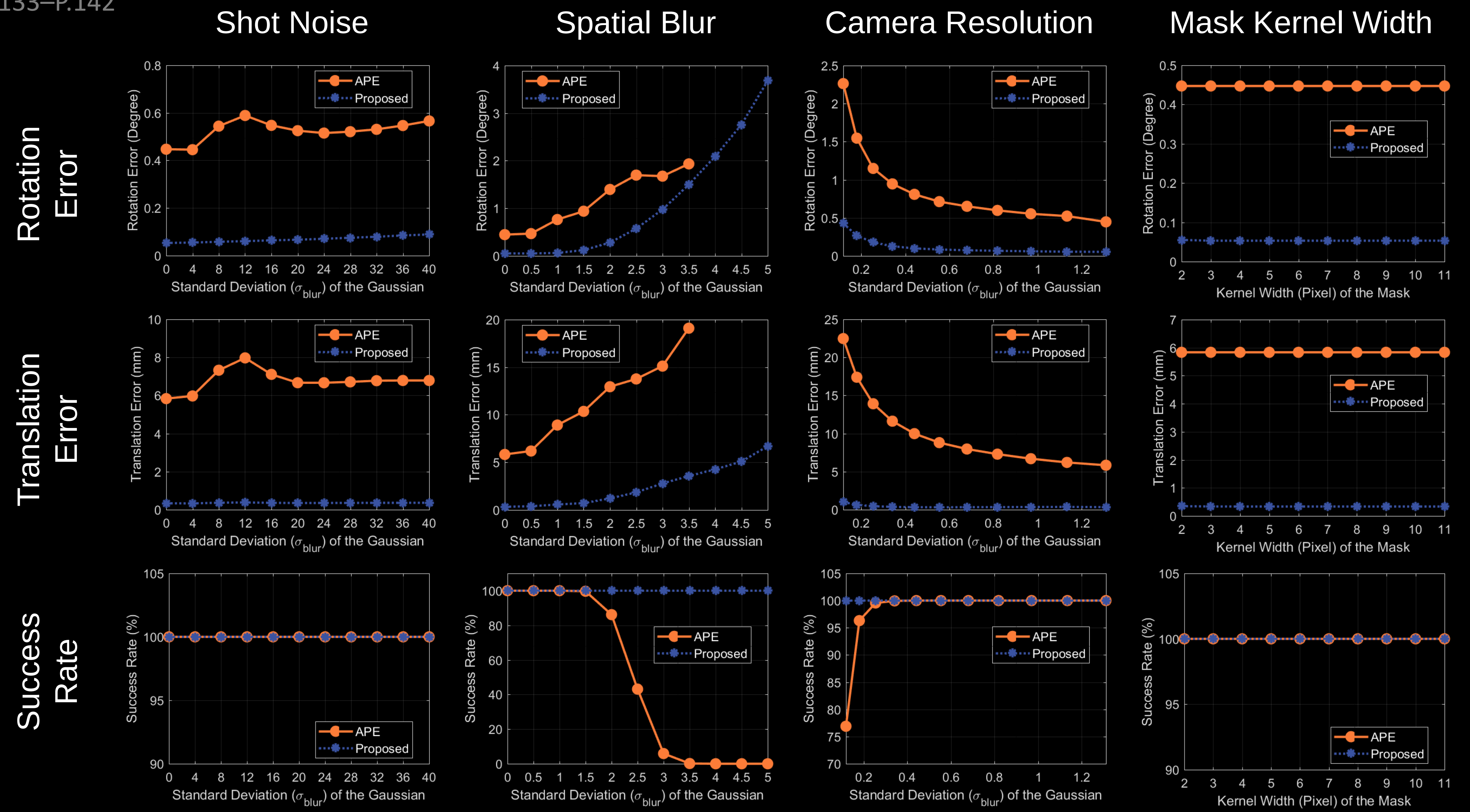
Evaluation with Synthetic Data

24 Motion Patterns

4 Varying Conditions




Pen-tip Trajectory 01
APE [5.074]
Proposed [0.244]
Ground truth
Y-Axis (mm)
80
100
120
140
160
180
-50
0
50
X-Axis (mm)
600
650
700
750
Z-Axis (mm)
Pen-tip Trajectory 02
APE [5.423]
Proposed [0.380]
Ground truth
Y-Axis (mm)
0
50
100
150
-200
-100
0
100
X-Axis (mm)
550
600
650
700
750
Z-Axis (mm)

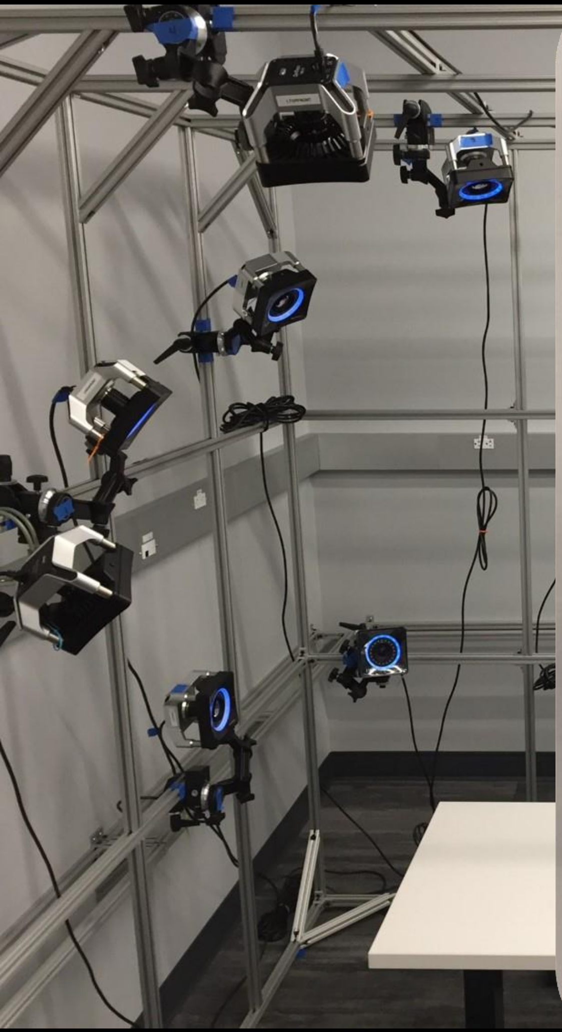


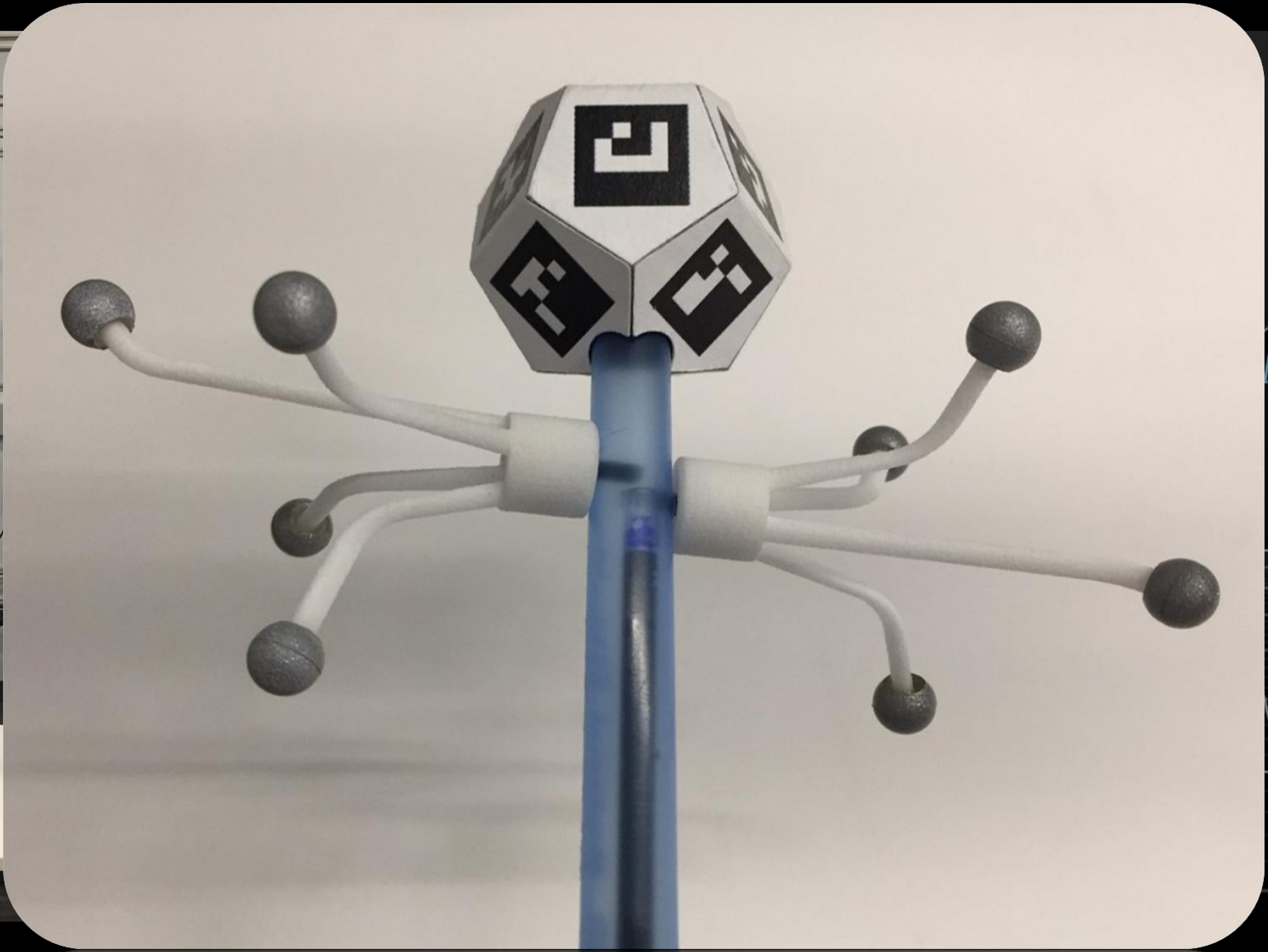
Shot Noise
Spatial Blur
Camera Resolution
Mask Kernel Width
Rotation Error
Translation Error
Success Rate
APE
Proposed
Rotation Error (Degree)
Translation Error (mm)
Success Rate (%)
Standard Deviation (σ_{blur}) of the Gaussian
Kernel Width (Pixel) of the Mask

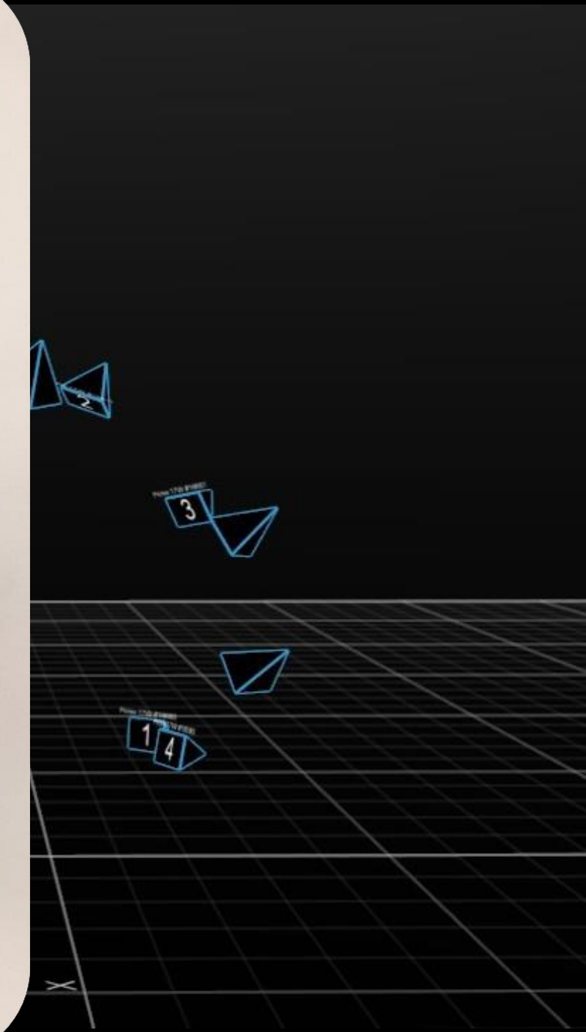
Evaluation with Real Data

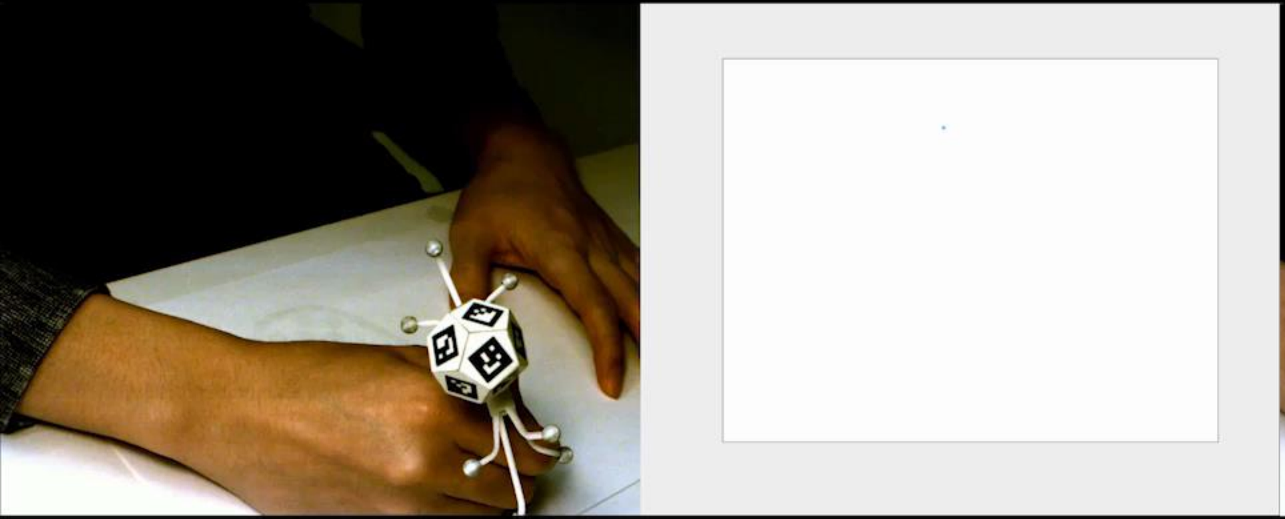
4 Real Drawings

VS. Mocap System (16 Cameras)

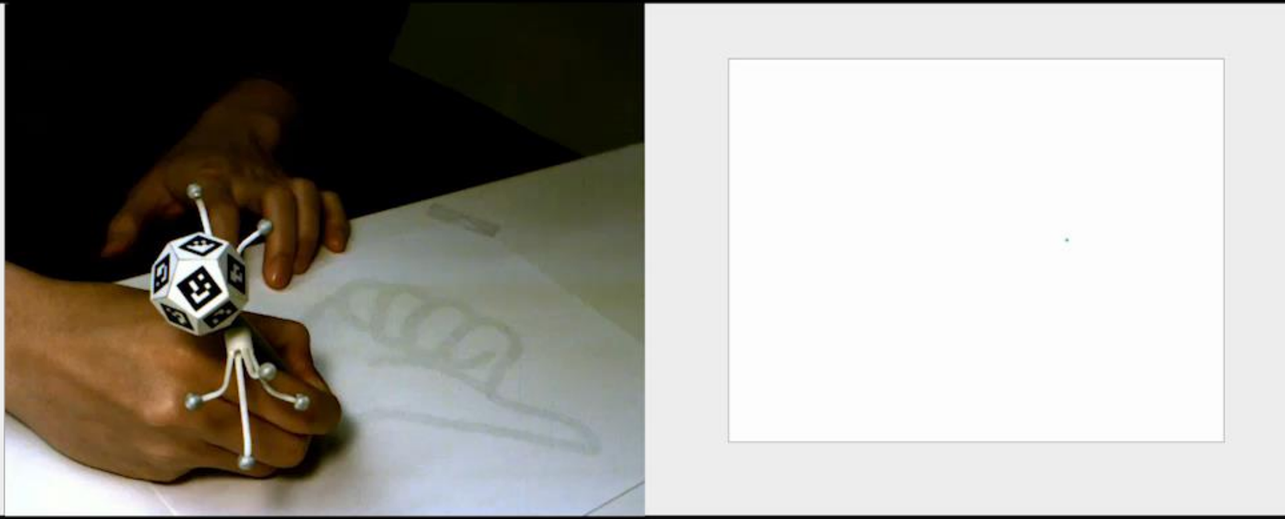




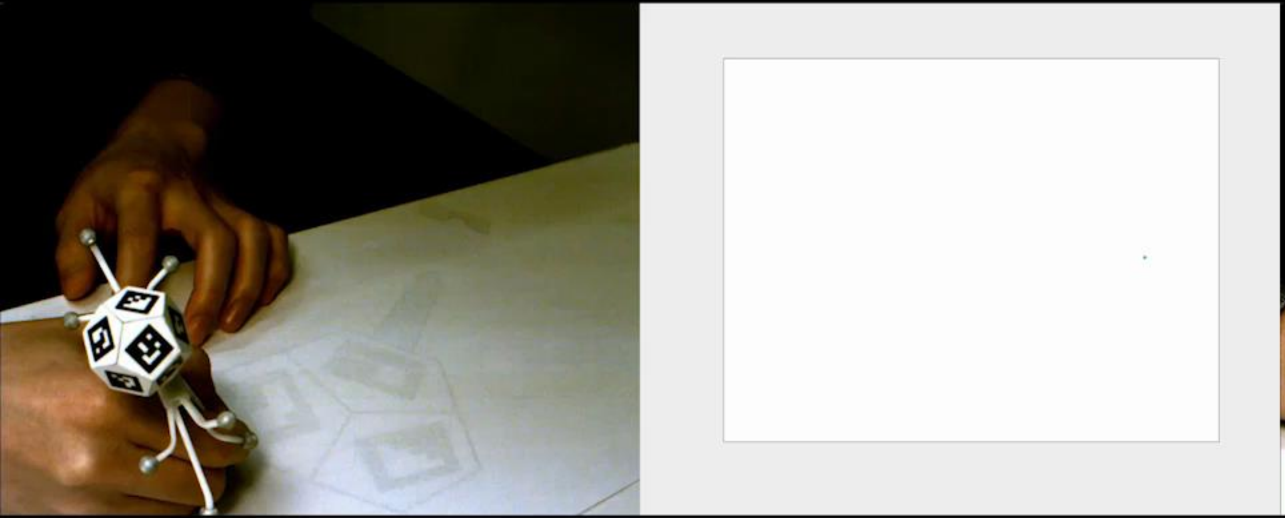

3
1
4



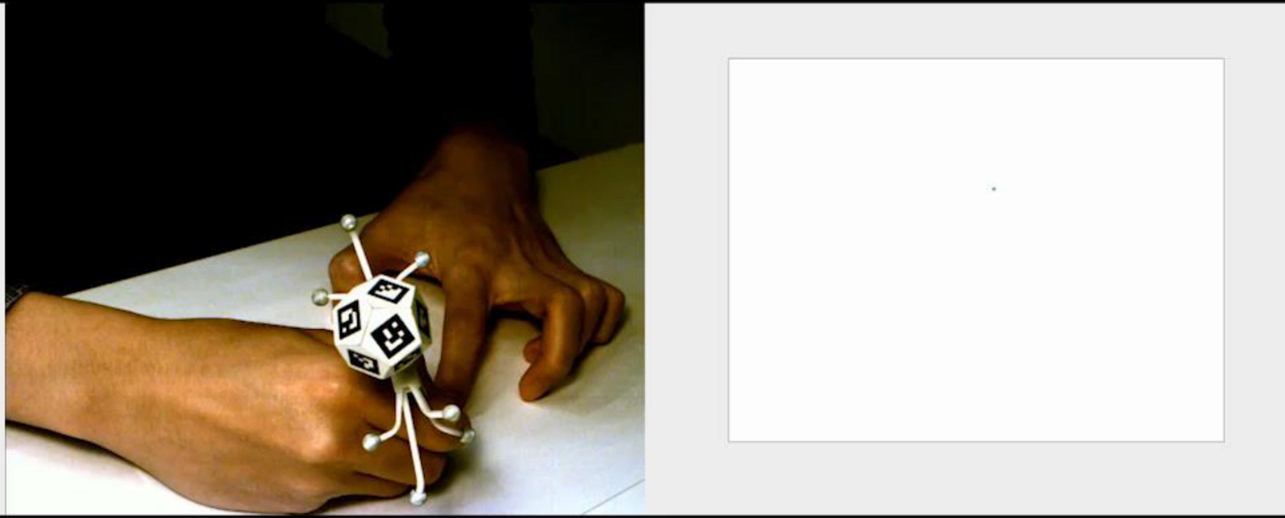
Boba



Thumb

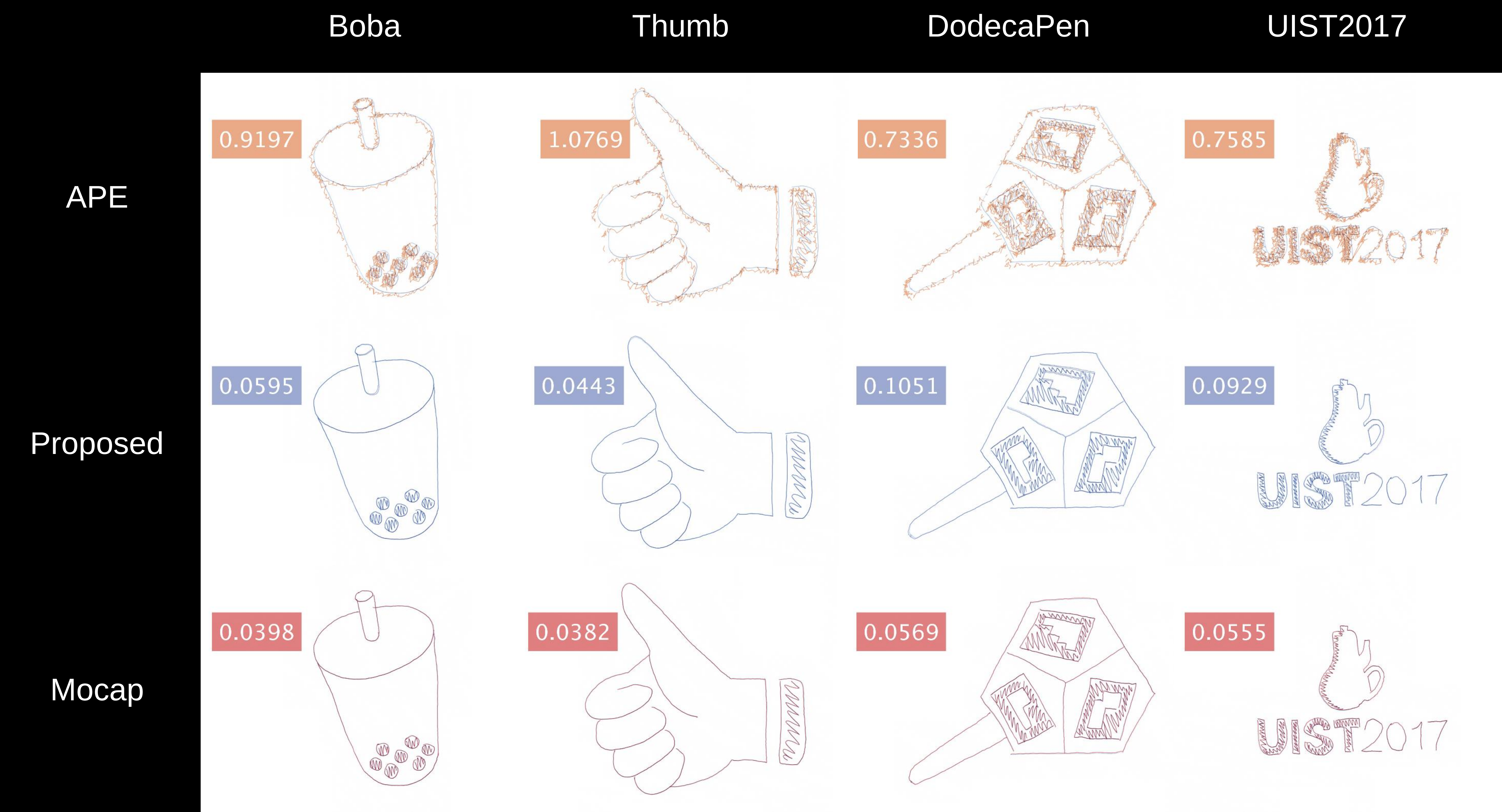


DodecaPen

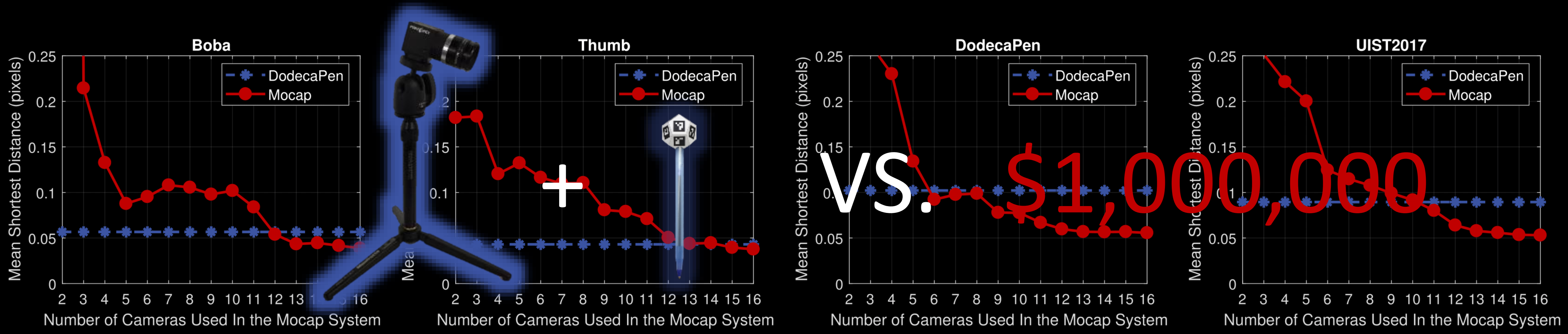


UIST2017

P.143–P.151


Boba
Thumb
DodecaPen
UIST2017
APE
0.9197
1.0769
0.7336
0.7585
UIST2017
Proposed
0.0595
0.0443
0.1051
0.0929
UIST2017
Mocap
0.0398
0.0382
0.0569
0.0555
UIST2017

DodecaPen VS. Mocap



Comparable to a mocap system with 10 active cameras

Main Achievements

1. OPT Dataset
   - A benchmark dataset for 6DoF object pose tracking
2. DPE Algorithm
   - A robust pose estimation method for planar objects
3. DodecaPen
   - A submillimeter-accurate 6DoF tracking solution

Future Work

- Learning-based pose estimation followed by dense pose refinement for general objects.
- Marker-based accurate 6DoF pose estimation and tracking solution.
- Pose recovering for planar objects using depth information and filtering techniques.


Thank You
For Your Attention ☺
DodecaPen
CLEAR
DELL