北京航空航天大學
BEIHANG UNIVERSITY

WORK SHARING

Indoor Localization for Quadrotors using Invisible Projected Tags

Jinjie Li, Liang Han*, Zhang Ren

International Conference on Robotics and Automation (ICRA) 2022

Jinjie LI

School of Automation Science and Electrical Engineering
Beihang University

March 12, 2022

SASEE
自动化科学与电气工程学院

北航中法工程师学院
Centrale Pékin

Overview

- Background
- Challenge
- IPT Method
- Experiment
- Future

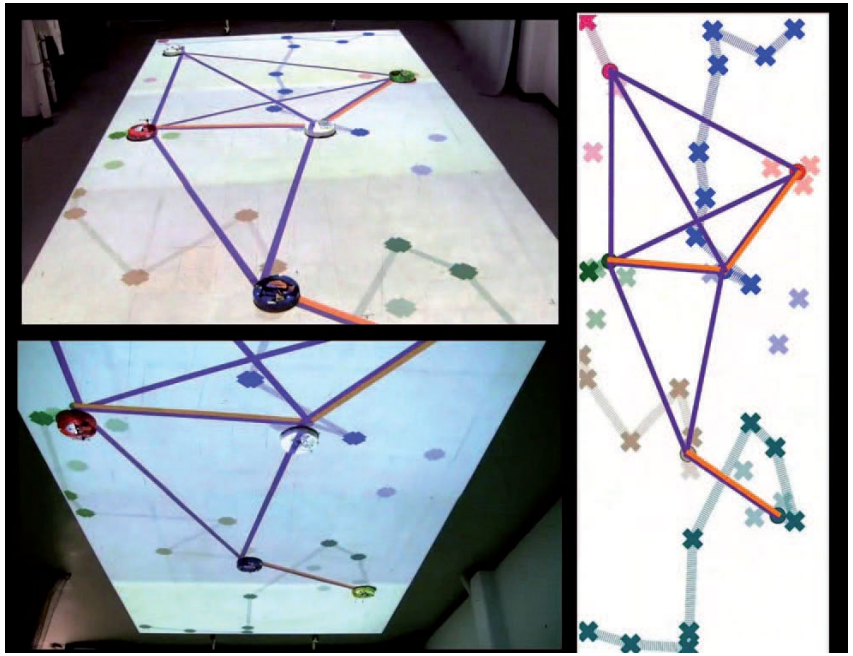
Background

- ❑ Difficult to show
- ❑ Hard to debug
- ❑ Separation of simulation and physical systems

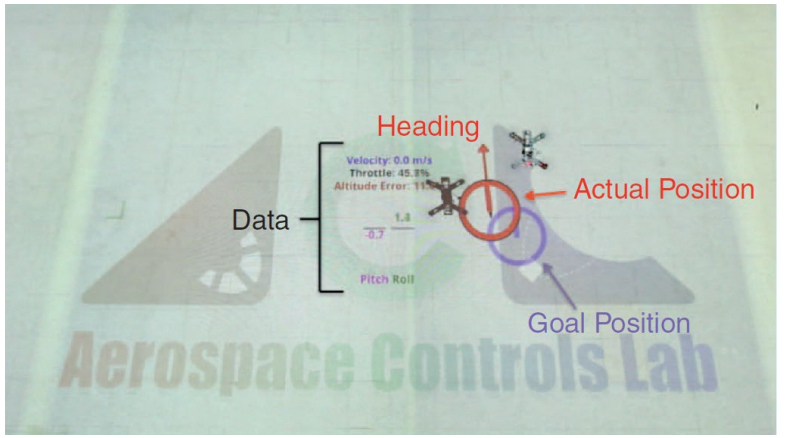
Challenge

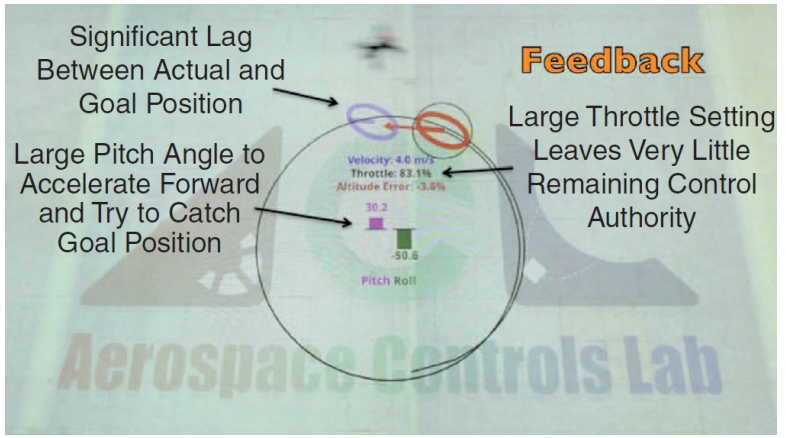
- Omidshafiei, Shayegan, et al. "Measurable augmented reality for prototyping cyberphysical systems: A robotics platform to aid the hardware prototyping and performance testing of algorithms." *IEEE Control Systems Magazine* 36.6 (2016): 65-87.

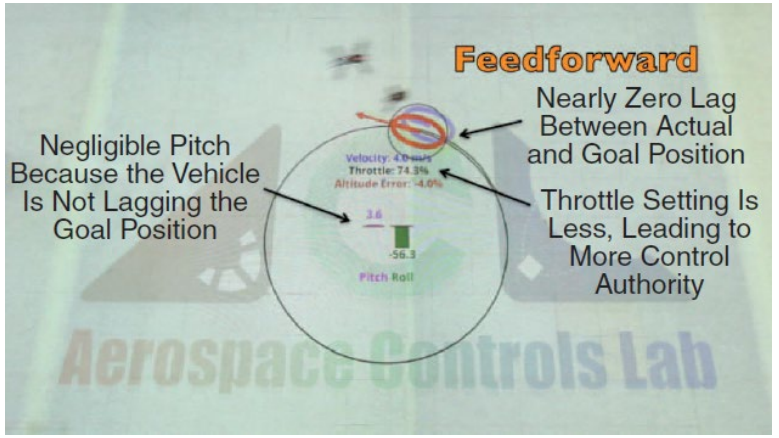
❑ Difficult to show

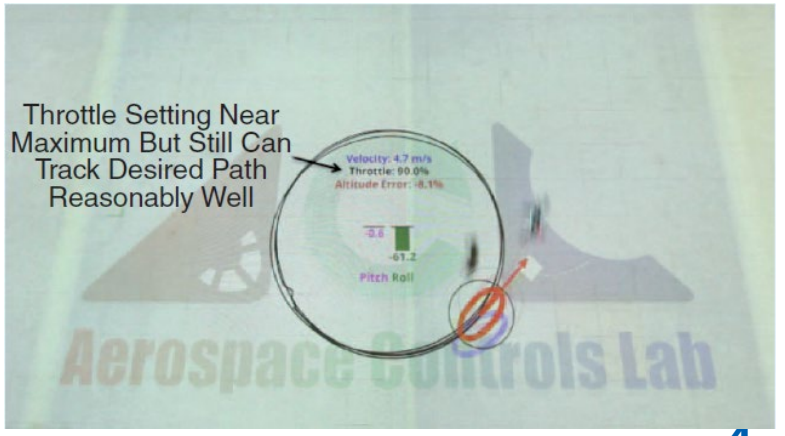


❑ Hard to debug





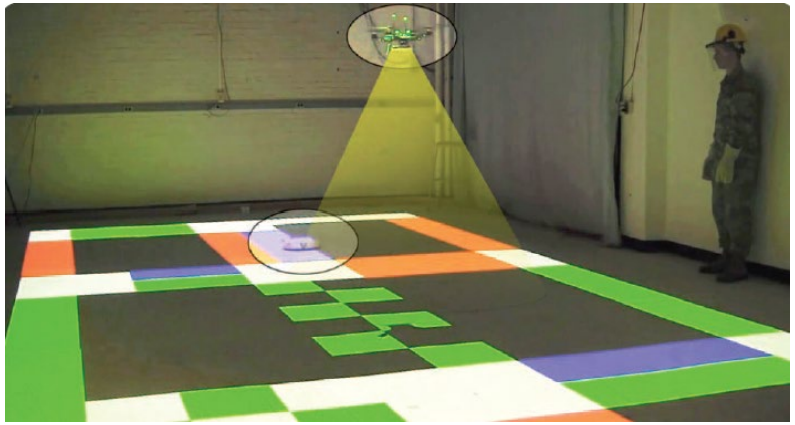


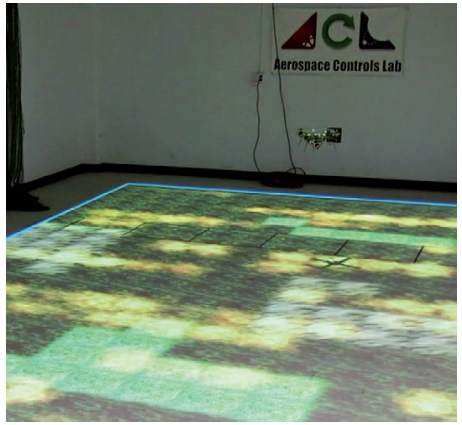


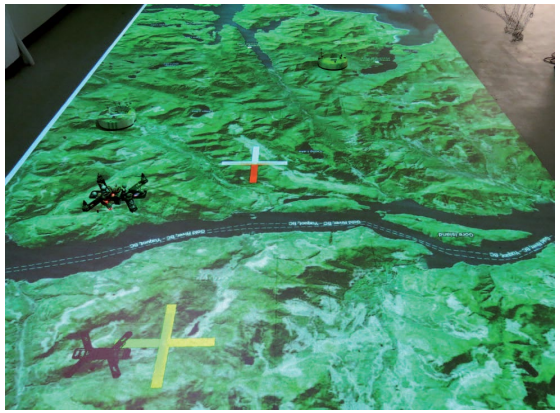
Challenge

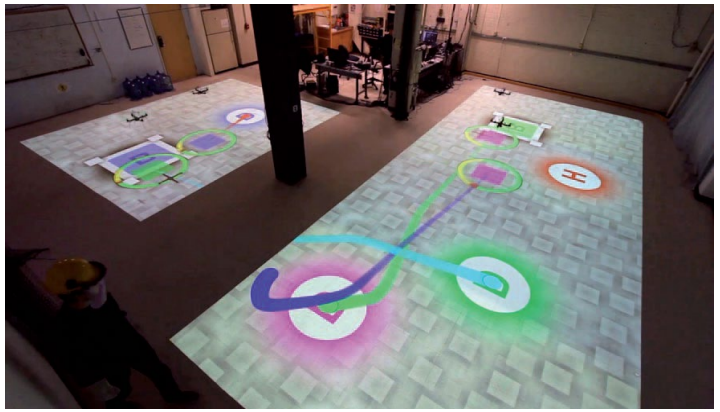
- Omidshafiei, Shayegan, et al. "Measurable augmented reality for prototyping cyberphysical systems: A robotics platform to aid the hardware prototyping and performance testing of algorithms." *IEEE Control Systems Magazine* 36.6 (2016): 65-87.

❑ Separation of simulation and physical systems

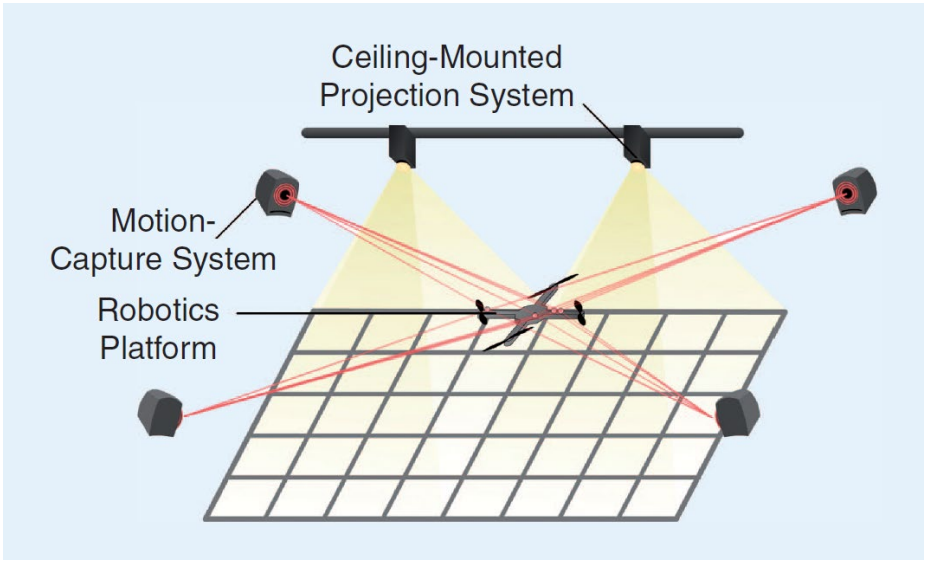








However…



Prime^{x} 13
$2,499

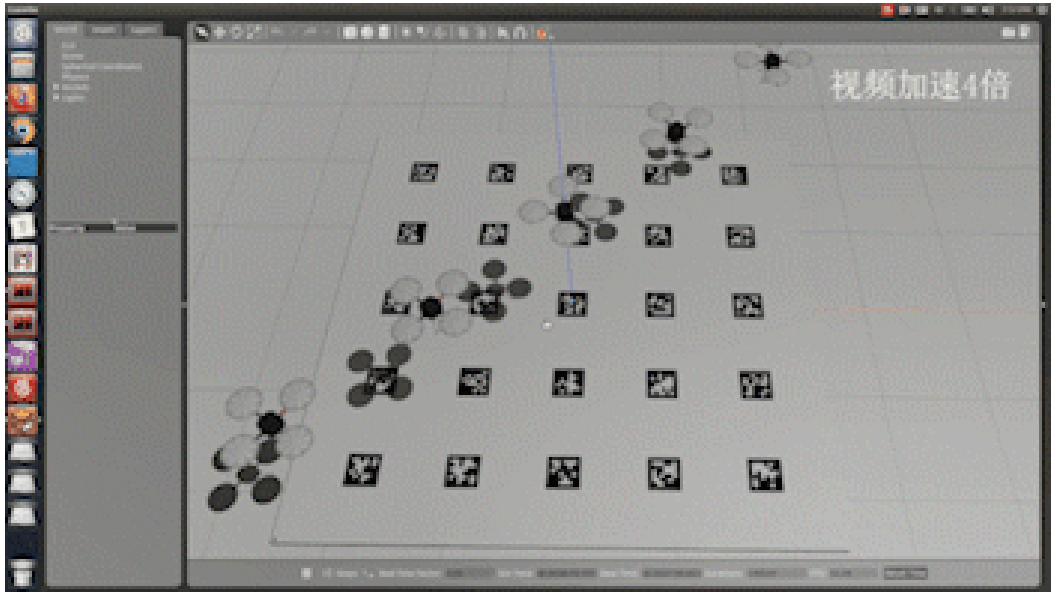
Prime^{x} 22
$3,999

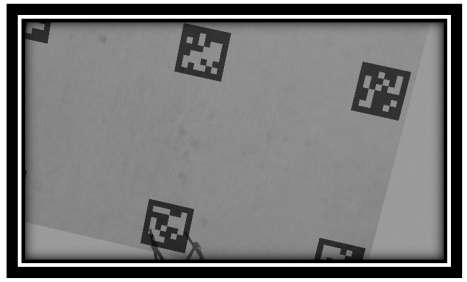
Prime^{x} 41
$6,499

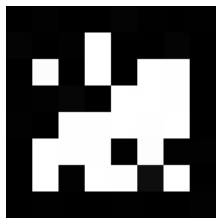
Motivation

本科毕业设计：

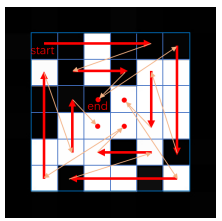
《基于视觉定位的多无人机编队控制方法研究》





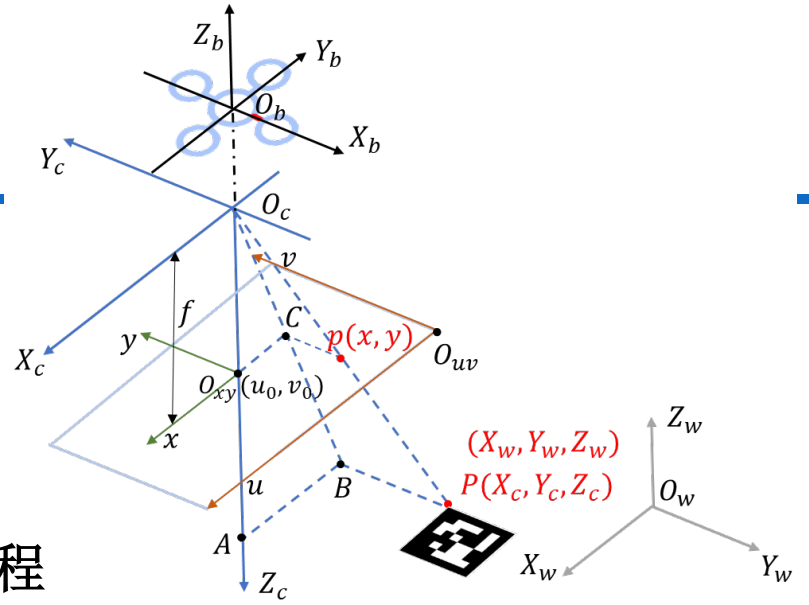


detection



decoding

0010,0010,0011,1011,1110,
1101,0111,1001,1101



视觉模型方程

像素坐标

摄像头外参数矩阵

$$Z_c\begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \begin{bmatrix} \frac{f}{dx} & 0 & u_0 & 0 \\ 0 & \frac{f}{dy} & v_0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} R & \vec{T} \\ \vec{0} & 1 \end{bmatrix} \begin{bmatrix} x_w \\ y_w \\ z_w \\ 1 \end{bmatrix} = M_1 \cdot M_2 \cdot X_w$$

世界坐标

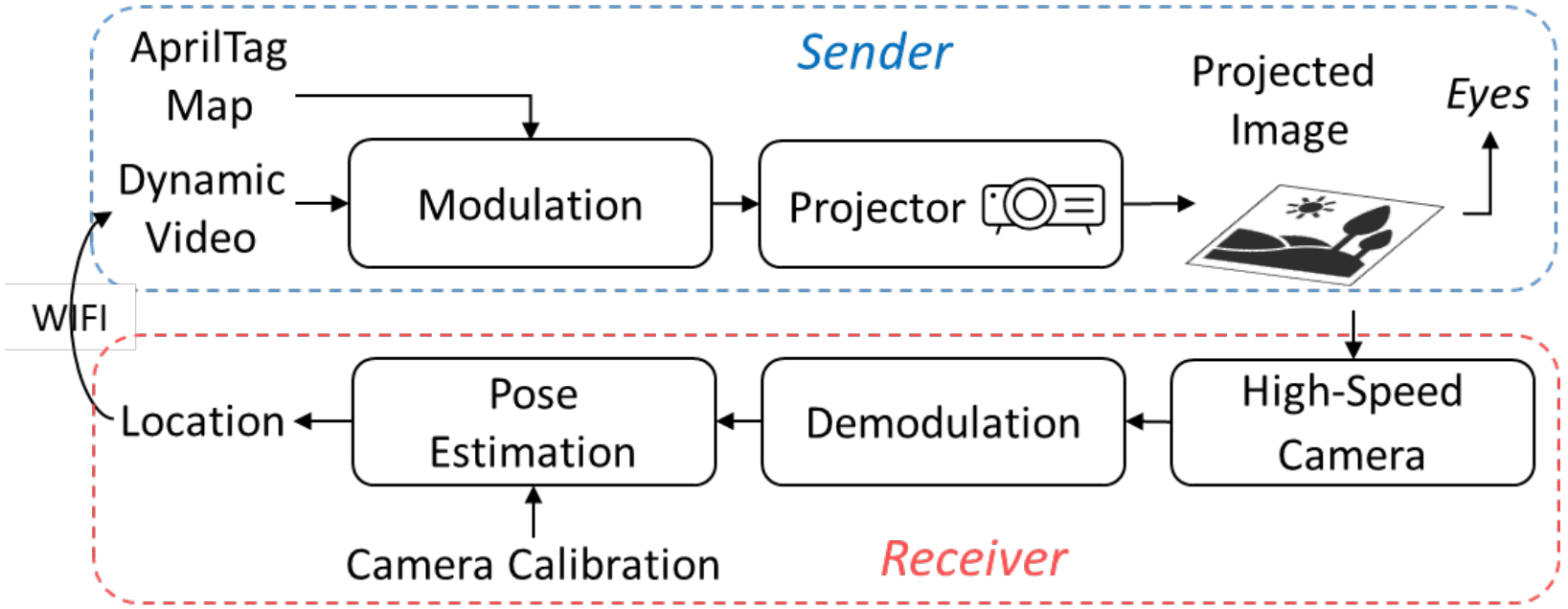
摄像头内参数矩阵

Screen-Camera Communication

IPT Method

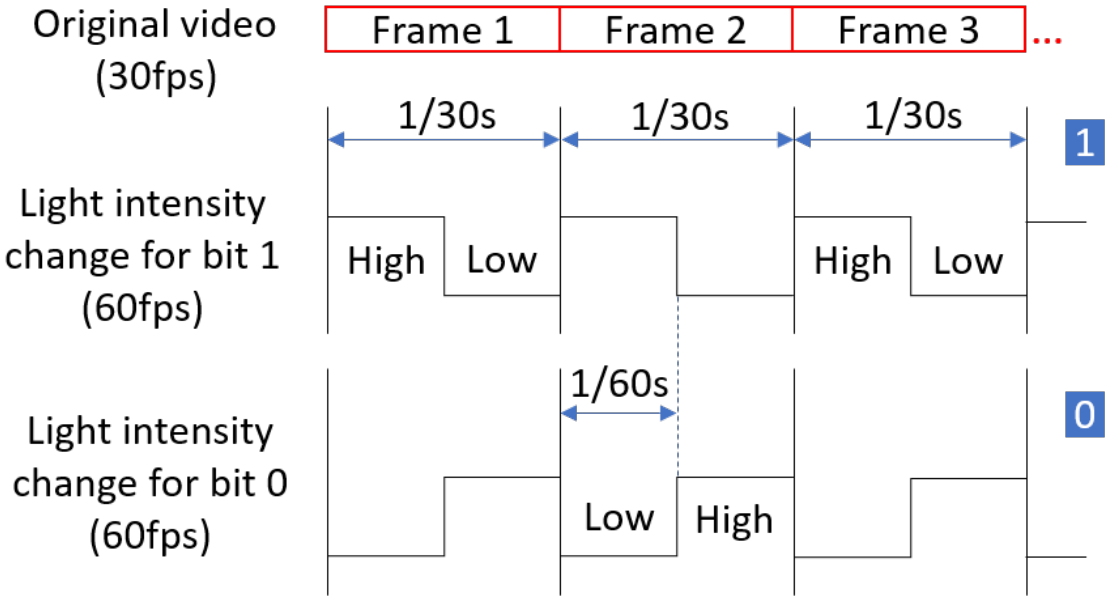
IPT

- Real-Time
- Centimeter-Level
- Low-Cost
- Quick-Start



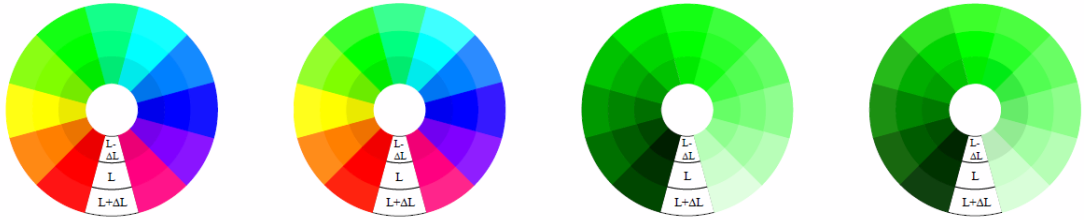
IPT Method

Sender-Modulation



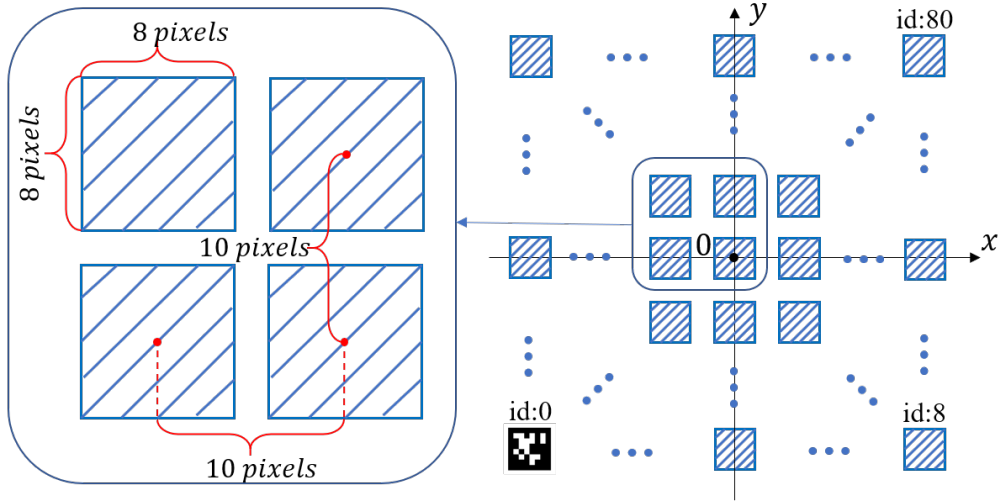
Encoding Process ($L \pm \Delta L$)

Zhang et al., 2021



Color Space Selection

$$\text{RGB} \rightarrow \text{CIELAB} \rightarrow \text{RGB}$$

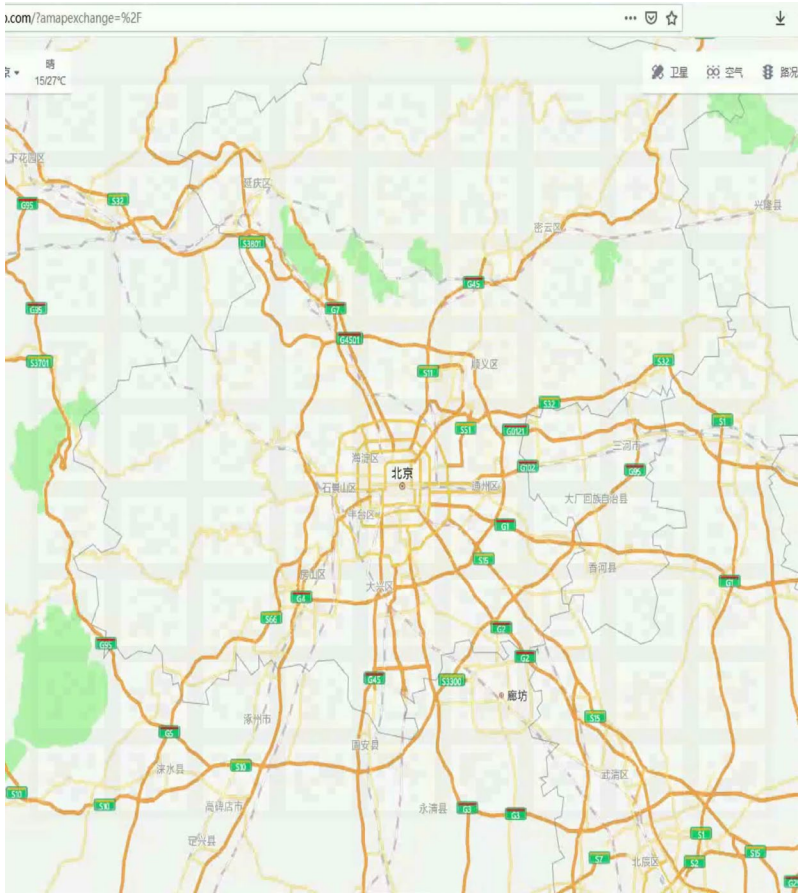


Tag Map Design

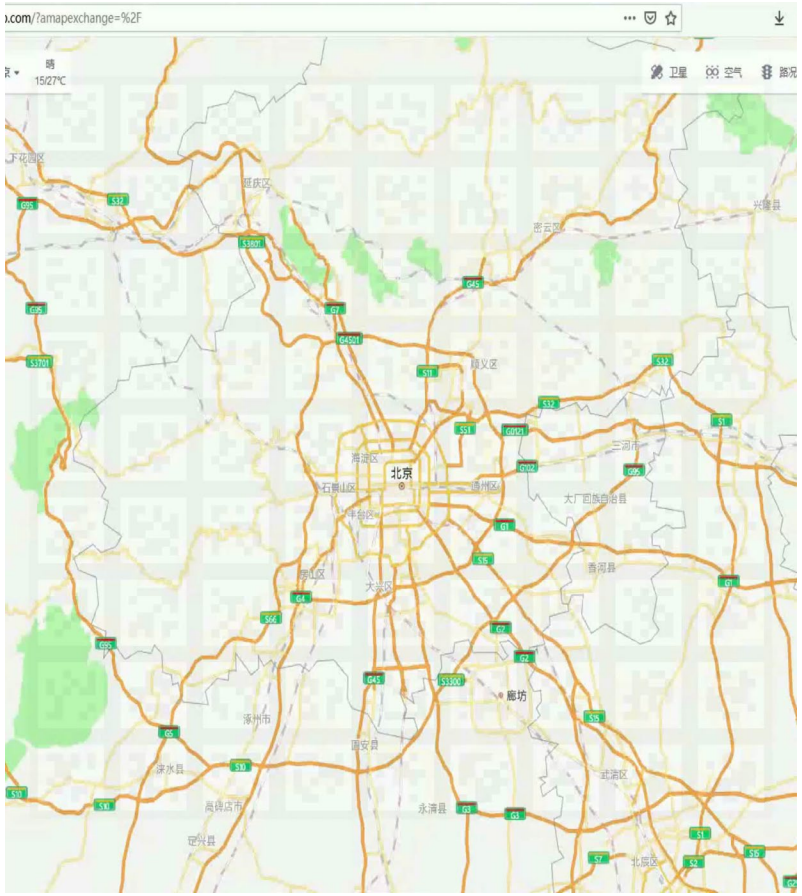
IPT Method

- The effect of ΔL

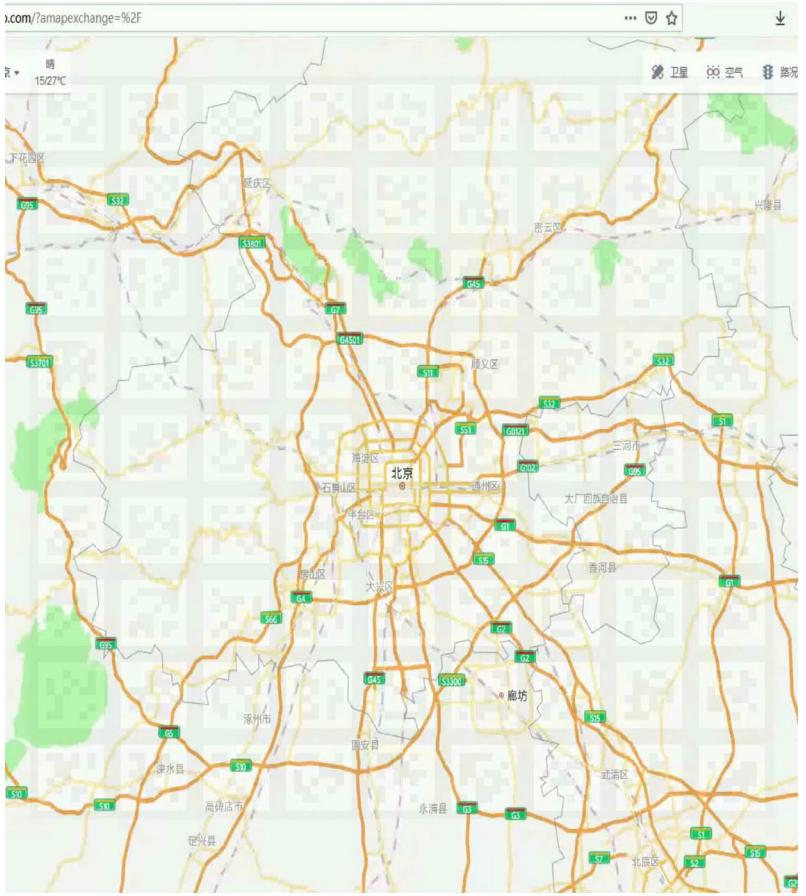
$\Delta L = 3$



$\Delta L = 4$

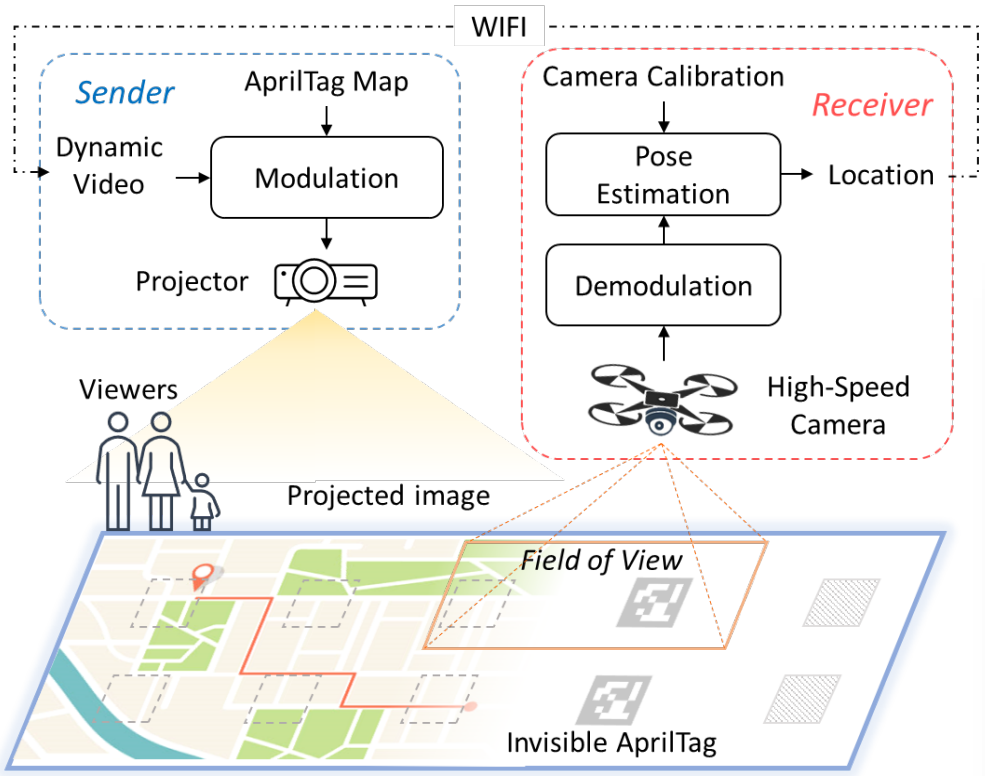


$\Delta L = 5$

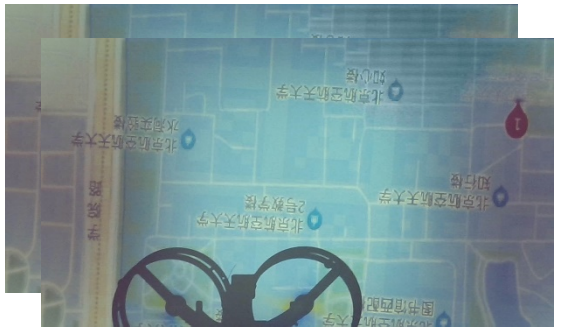


IPT Method

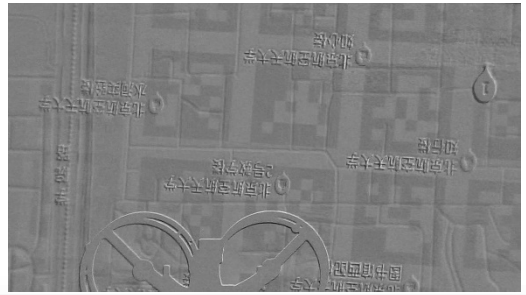
Receiver-Demodulation

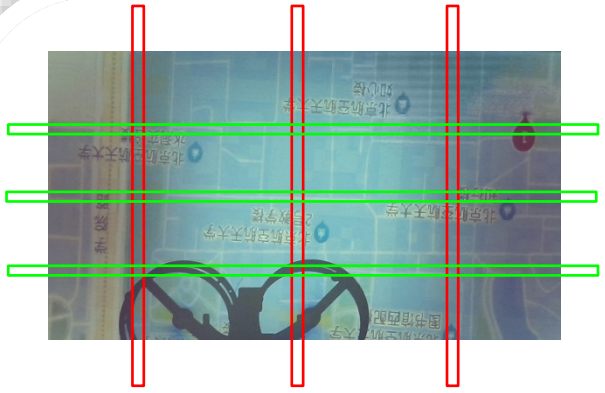


Two successive frames

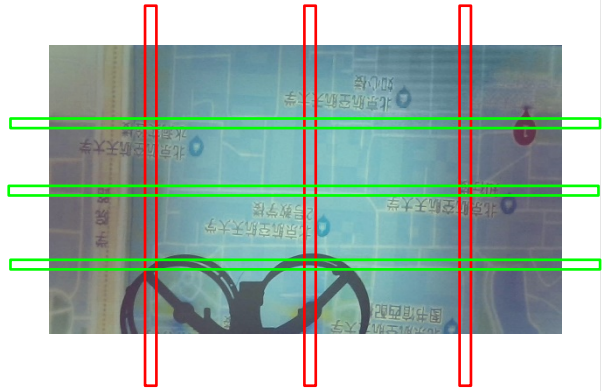


Subtract in lightness channel





Frame 1

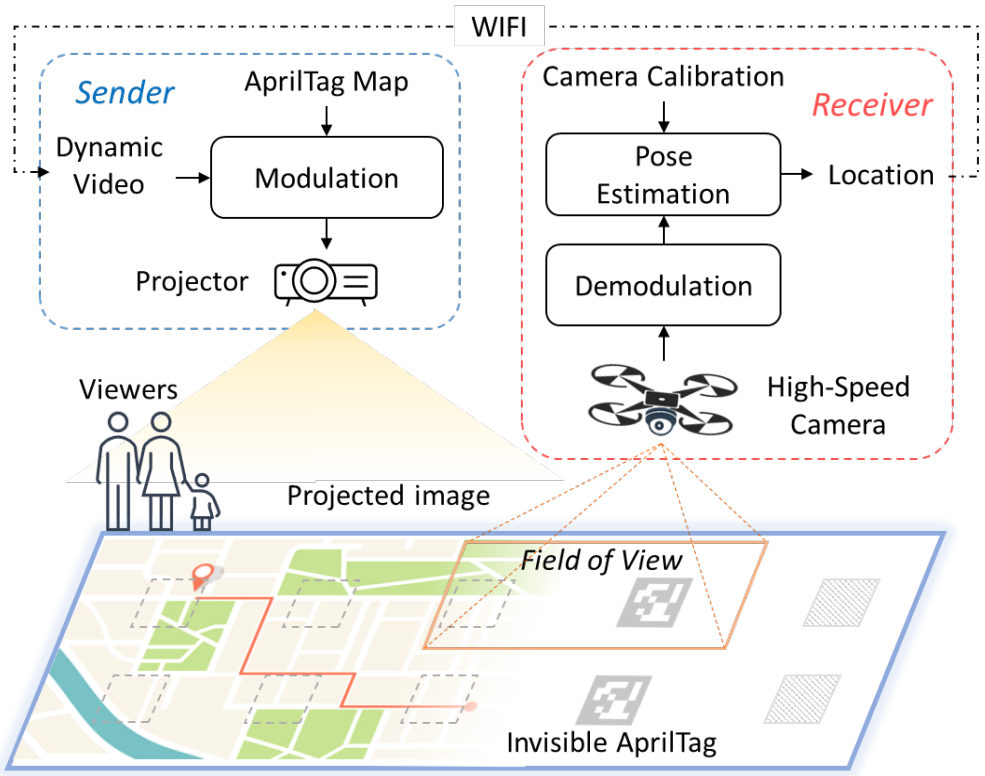


Frame 2

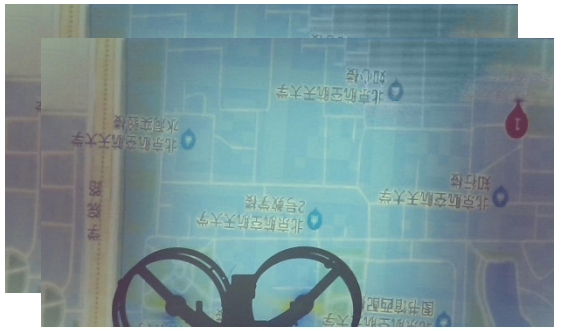
$$m_y^* = \arg\min_{|m_y| \le b} \begin{cases} \sum_{i=1}^{N} \|T_i^{pre}(0:h-m_y) - T_i^{now}(m_y:h)\|_1 \\ \qquad (m_y \ge 0), \\ \sum_{i=1}^{N} \|T_i^{pre}(-m_y:h) - T_i^{now}(0:h+m_y)\|_1 \\ \qquad (m_y < 0) \end{cases}$$

IPT Method

Receiver-Demodulation



Two successive frames



Subtract in lightness channel

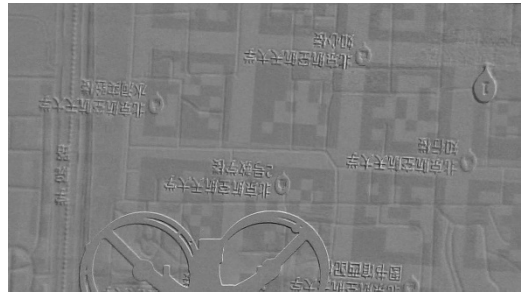
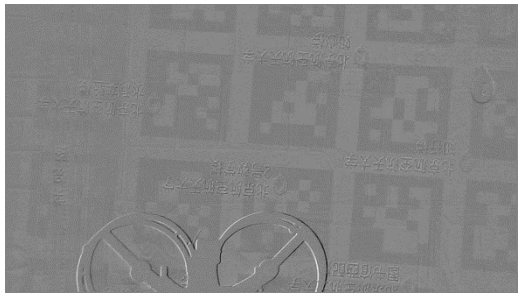
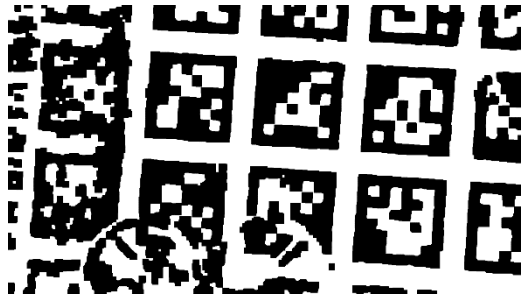


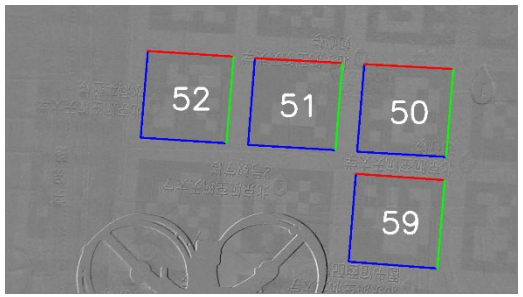
Image alignment



Preprocessing

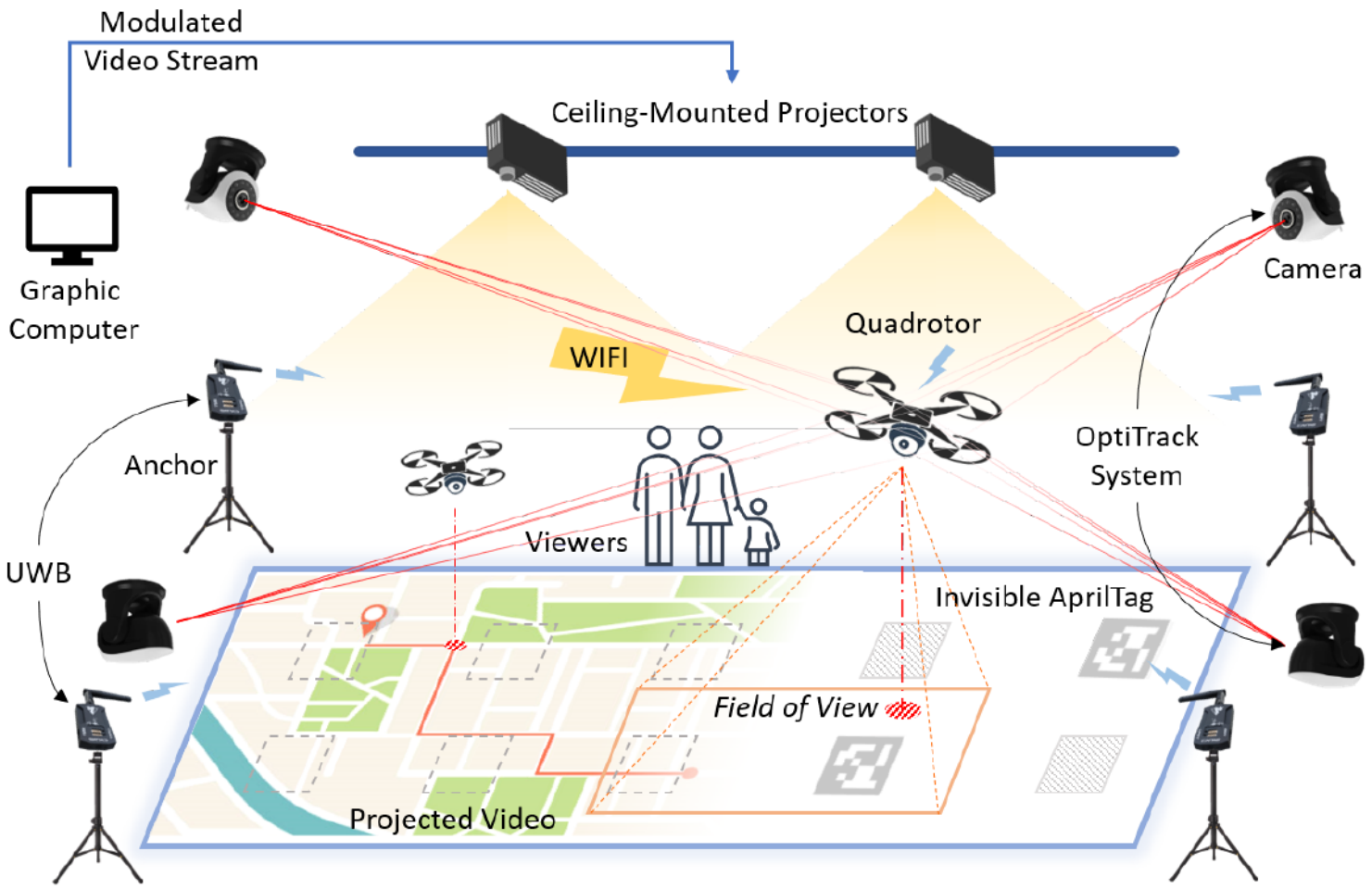


Tag detection



Pose Estimation

Experiments

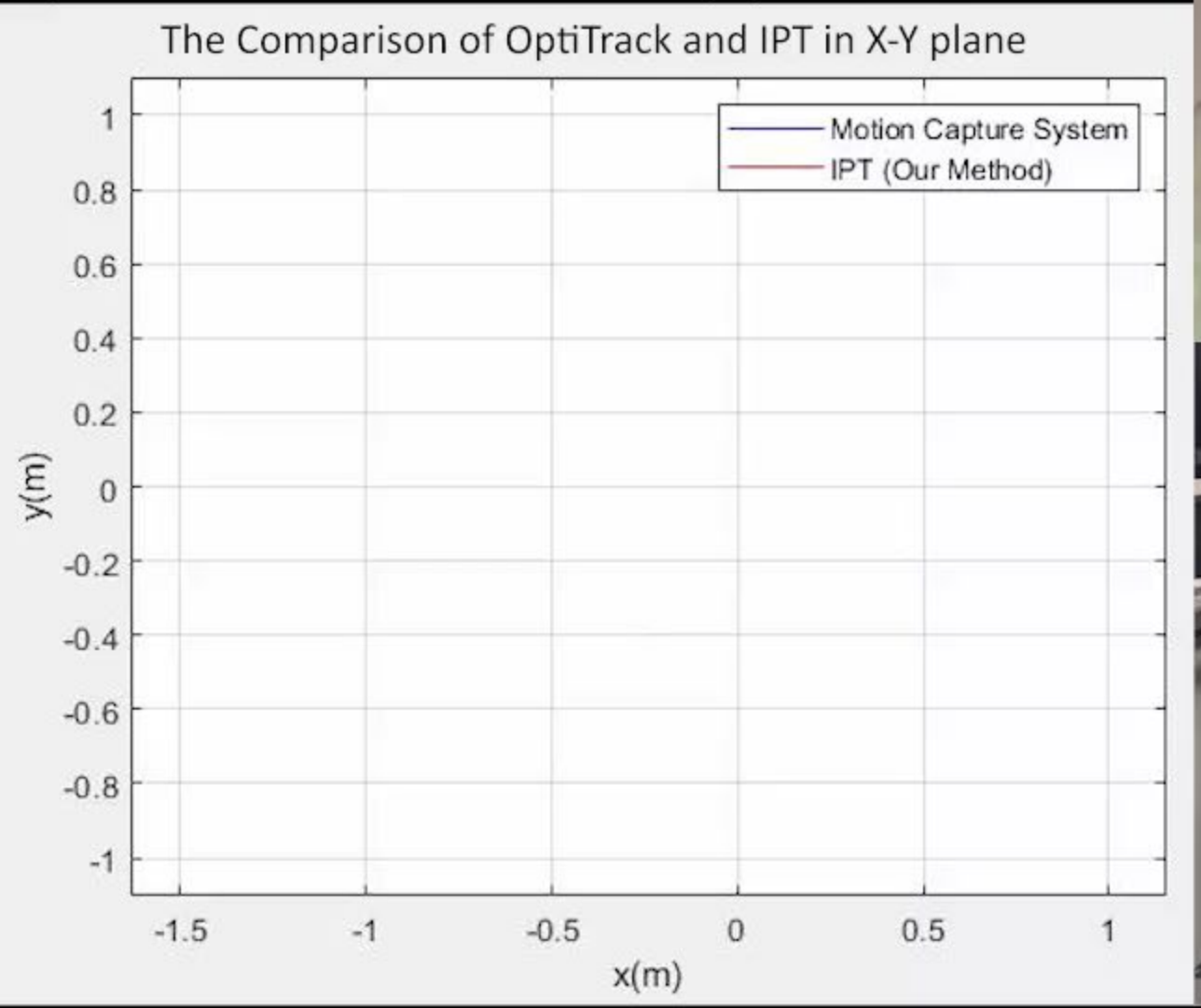


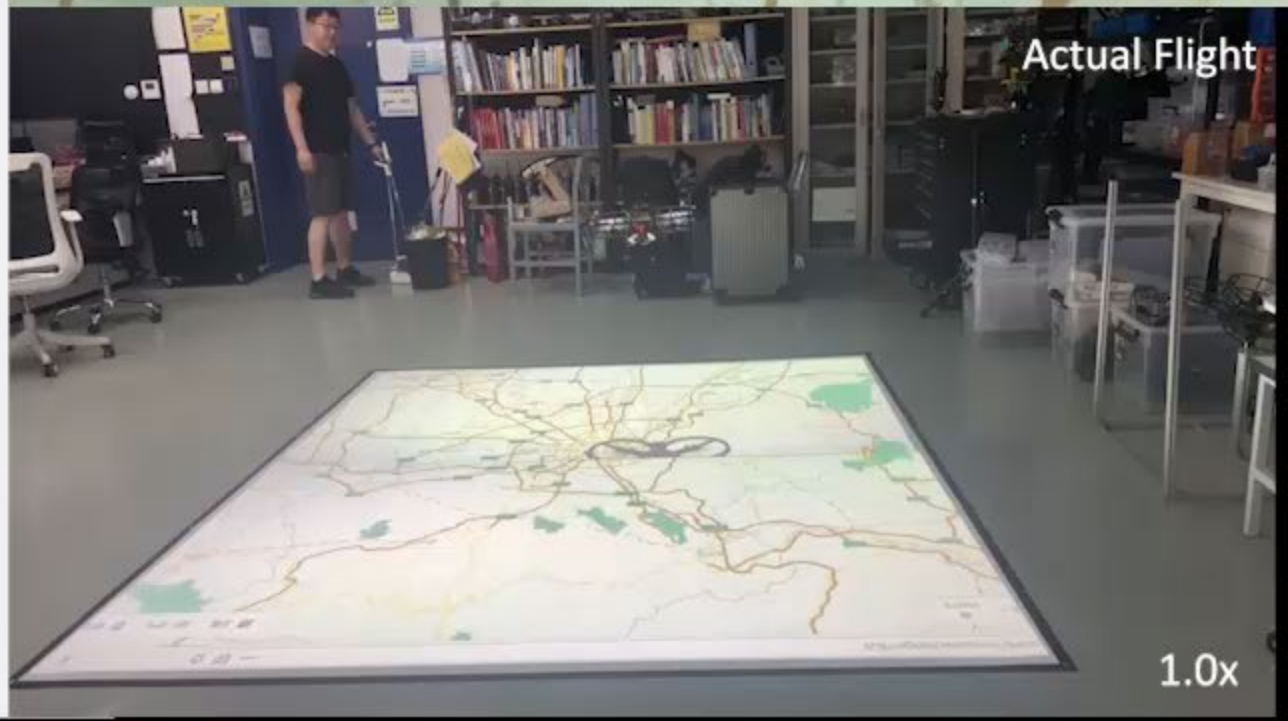
Experimental Setup

- IPT, UWB, OptiTrack
- Human pilot, rectangle, 0.8m height
- A 15-second video as the test unit
- First stage: post-processing for accuracy
- Second stage: real-time for speed

Accuracy experiment:

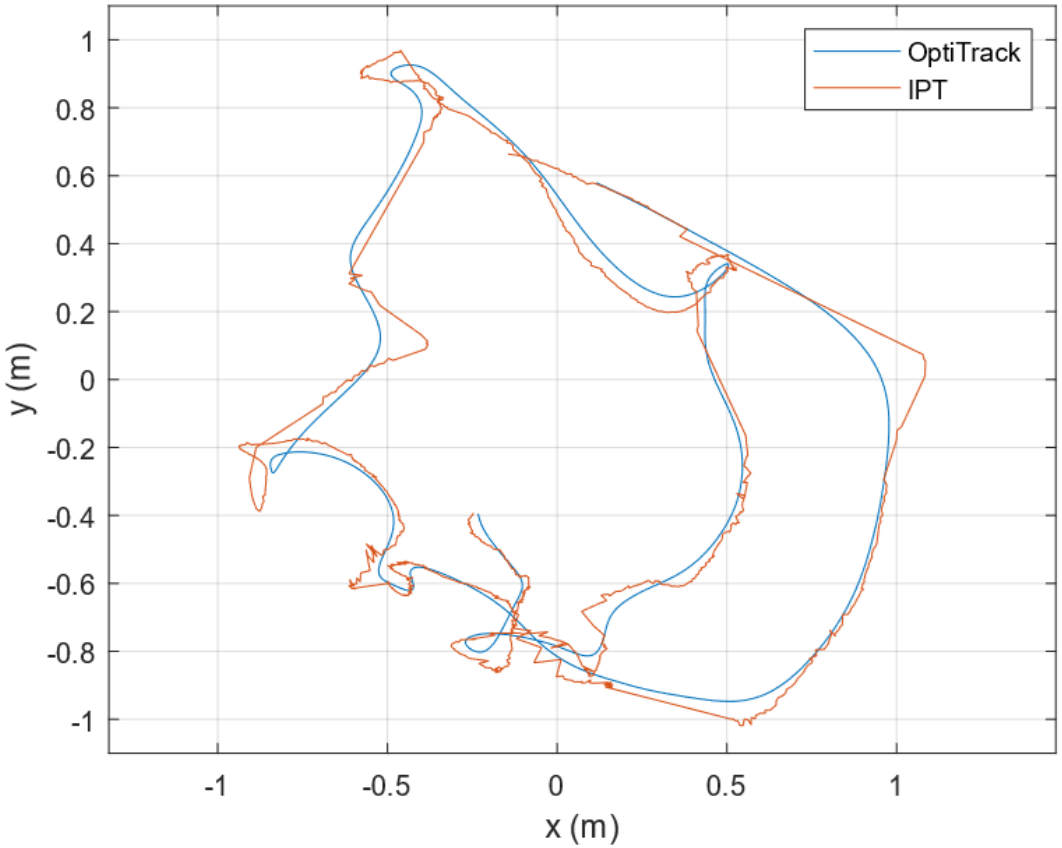
*Note that this result comes directly from OpenCV function SolvePnP(). The final result can be obtained after inversely coordinate transformation, which is significantly affected by orientation measurements. Read paper for details.

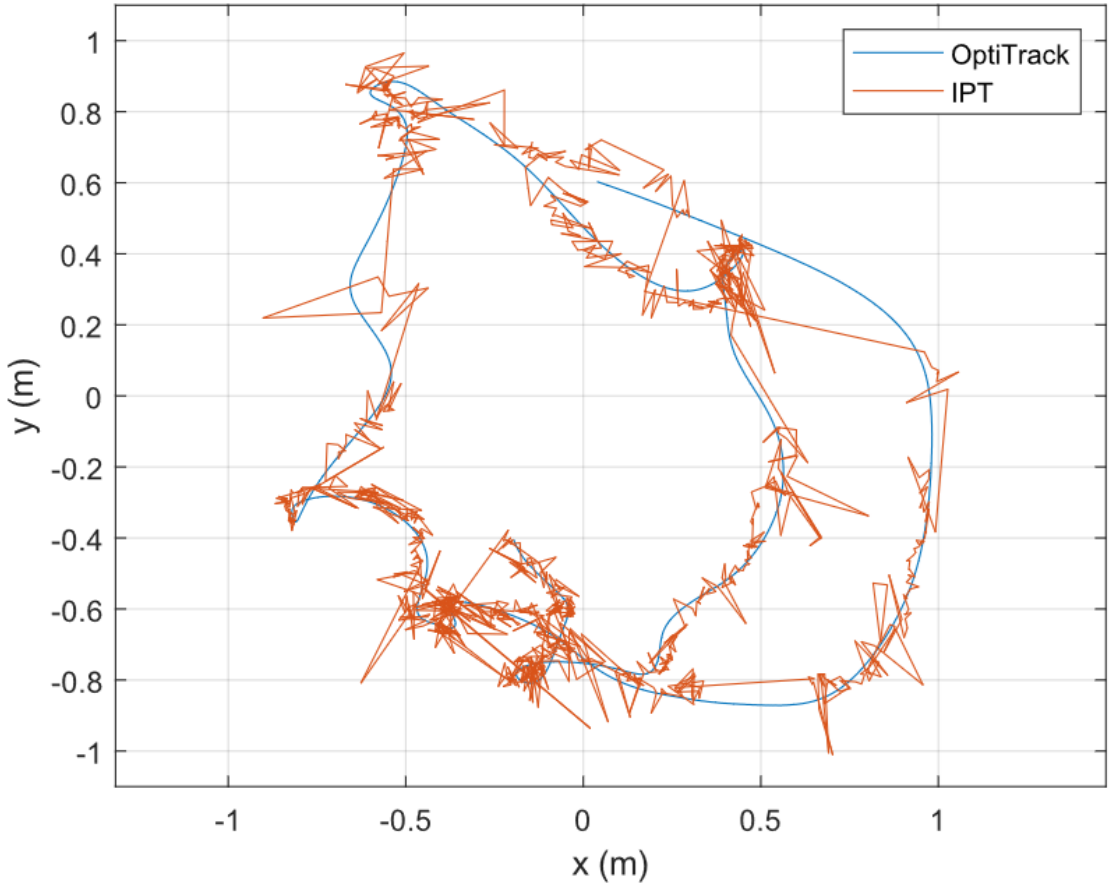




Experiments

Analysis

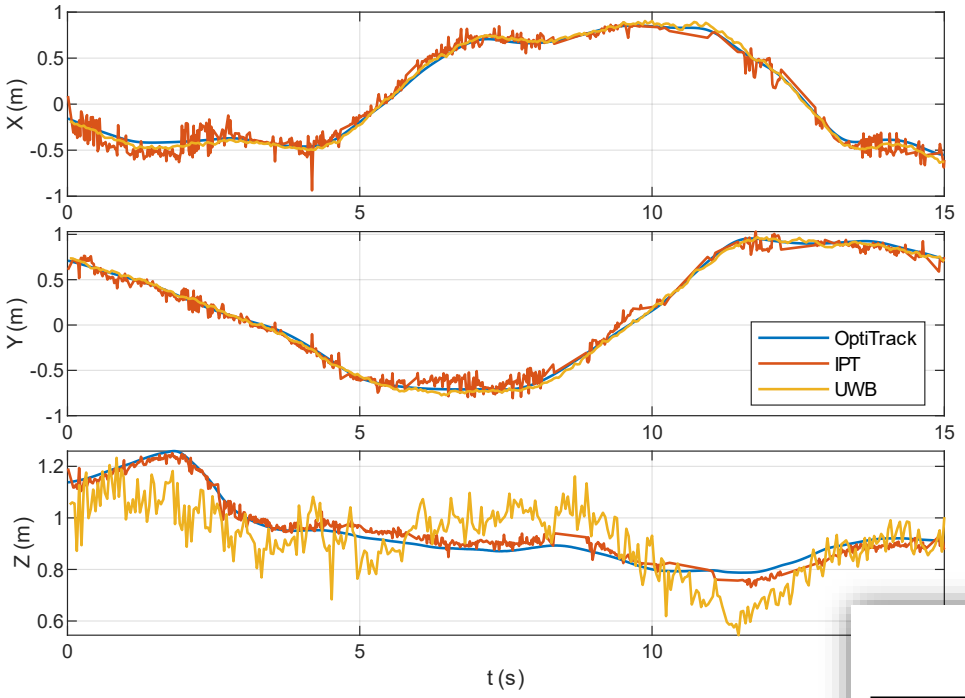




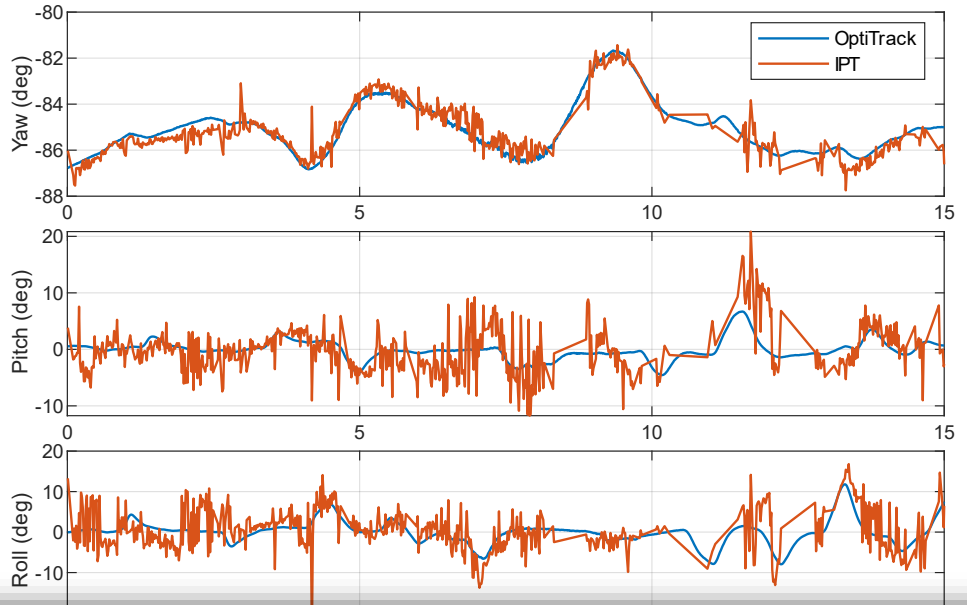
$$\left[\begin{array}{c|c} {}^{w}_{c}R & {}^{w}\vec{T} \end{array} \right] = \left[\begin{array}{c|c} R^T & -R^T \cdot \vec{T} \end{array} \right]$$

Experiments

Position



Orientation



POSITIONAL ERROR AND ROTATIONAL ERROR (**METRICS**: MAE

	X Error [m]	Y Error [m]	Z Error
UWB	**0.0305 / 0.0354**	**0.0232 / 0.0251**	0.0943 / 0.
IPT	0.0690 / 0.0886	0.0535 / 0.0674	**0.0230 / 0.**

	Yaw Error [rad]	Pitch Error [rad]	Roll Error
UWB	—	—	—
IPT	**0.352 / 0.460**	**2.659 / 3.657**	**3.342 / 4.**

THE COMPARISON OF THREE LOCALIZATION METHODS.

	OptiTrack	UWB	IPT
Price[1]	CNY 190,000	CNY 35,000	**CNY 30,000**
Position Accuracy	sub-mm level	cm level	cm level
Orientation Info	YES	NO	**YES**
Anti-Interference	YES	NO	**YES**
No Calibration	NO	NO	**YES**

1.0x

Conclusion

- IPT
  - We proposed IPT, an indoor localization method based on human-invisible projected fiducial tags. This method utilized the visual characteristics of human eyes to hide markers.
  - To test the performance, we designed an AR quadrotor flight platform, of which the structure consists of the sender part and receiver part.
  - Indoor flight experiments were conducted to evaluate accuracy and speed of IPT. The result presented centimeter-level accuracy and a frequency of ten FPS.
  - The economical and quick-start features made IPT an appropriate solution for AR robotics platforms.

- A balance of **existing knowledge**, **difficulty**, **laboratory equipment**, and **luck**.

Conclusion

Reviewers' Comments

Reviewer 1

The idea is quite clever and the authors build a working system on top of it in a fairly straight-forward way, dealing with a few practical difficulties. Specifically, the authors extract the image of tags by subtracting two successive images from the projector, however motion of the

All in all, the cleverness of the proposed idea scores high marks for me, while I would encourage the authors to strengthen the paper by addressing some of the suggestions above for the final submission.

Reviewer 3

The novelty seems to be on the Modulation/Demodulation, which is responsible for embedding and recovering the tag from the lightness channel. However, is it the first time this "screen-camera communication" method has been used like this? The reader cannot be sure simply from the cited references. The AR projection and pose estimation have been extensively used before, and there are no contributions in these parts.

Reviewer 2

17501

The authors presented a framework for indoor localization for quadrotors using a camera and projectors. The differential of this work is the convenience of its use in augmented reality applications using projectors. The projectors, fixed to the ceiling, project a video in which the frames have been modified to encode a map of fiducial markers in such a way that it is invisible to human eyes. A downfacing camera, fixed to the drone, captures the projected images and decodes them, extracting the markers and tracking the drone position using PnP.

In general, the idea is interesting, and the concepts presented in the paper are clear, but I recommend improving these points:

== Major ==

- The limitation of the method should be better discussed: minimal and maximal height; ambient lighting changing; dynamic video or complex videos; multi robots; detection area or volume; drones take-off and landing;

- Only accuracy in the position of the method was assessed. Why did you not assess orientation? Noisy? How much?;

- By assessing orientation, this is also a differential regarding the UWB technology;

- The proposed method is convenient for augmented reality applications, but you do not assess how well the result is (Did people notice the markers on the projection?);

- "For non-cooperative localization technology such as simultaneous localization and mapping (SLAM), heavy sensors enlarge the drone frame, making these drones impossible to be used in an ordinary room." not necessary! SLAM can also be visual-based.

- "Through WIFI" which protocol? TCP-IP? UDP?

- You should define tag as a visual fiducial marker at the beginning of Section III;

- Fig. 3: Explain that 0/1 comes from the marker map and concerns the 2D coordinates of the image. This becomes clear only after I have seen algorithm 1.

- How C_map, the configuration of the AprilTag map is created? It sounds to me that it requires measuring the project markers on the floor concerning a referential frame. Is a calibration step required or not? Supposed I changed the projector height (relative to ground), the markers became smaller.

- "From the table, the average positioning accuracy (mean Average Precision, mAP)"... precision and accuracy are not the same things.

- How is the localization method expected to behave with sudden movements in the drone's orientation? When pitch and roll speeds are higher, two consecutive images are somewhat farther apart, making alignment and decoding difficult. You only evaluated the speeds on the X and Y axes.

- The projected video looks somewhat well-behaved; do you expect similar results for more complex videos?

- Tip: What if your method dynamically changes the size of the markers depending on the estimated height of the drone?

== Minor ==

- Fig. 1 is not referred to in the text;

- I suggest adopting human-invisible markers instead of only
invisible;

- Check writing: air quotes should be pair of quotations; et
at should be italic;

- "The motivation is to design an inexpensive positioning method for indoor AR robotics systems, and hence the tag-based approach is chosen in this work." This should be part of the introduction section for better organization of the text;

- "but the (very) high price limits";

- "For indoor positioning, the information transmitted is relatively (little), and hence the (algorithms that have good invisibility) and reliability" improve this sentence;

- "Fig. 4 shows the position information to embed, an image." improve.

- Where is the world referential frame for IPT?

Future Work

Next Generation Educational Robotics Platform: A Moveable Augmented Reality Approach

Jinjie Li, {Xuan Cao, Jingyi Huang, …}, Liang Han*, Zhang Ren

➢ RA-L或CSM等包罗万象的期刊。

- Synchronize problem
- The impact of the specific steps
- Consider the rotation movement
- Error analysis
- Ground vehicle
- Multi-agent
- ……

Thanks!

Q&A