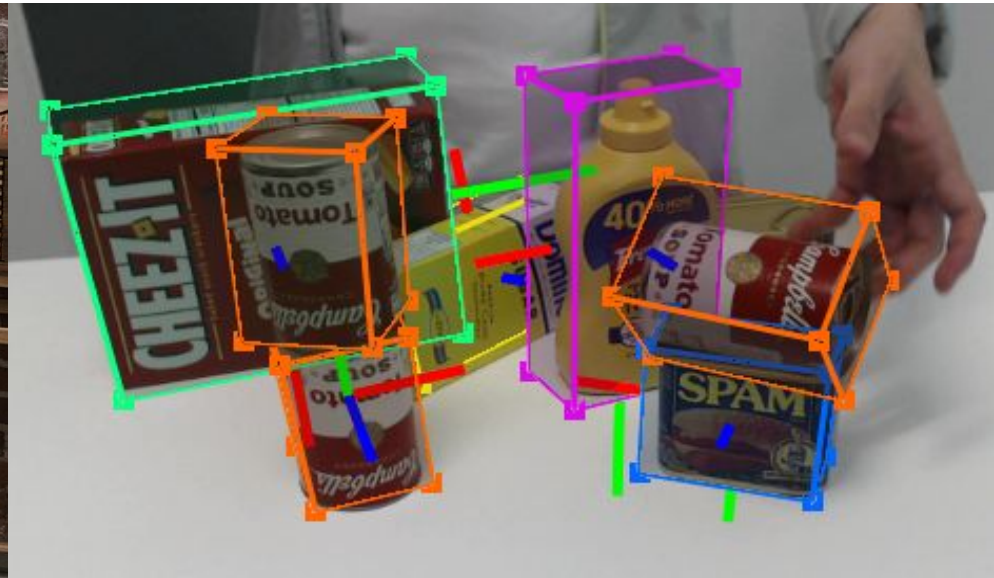
Depth aware Real-time 6DoF Object Pose Tracking for Augmented Reality

Student ID : 20213213
Name : Chaejung Maeng

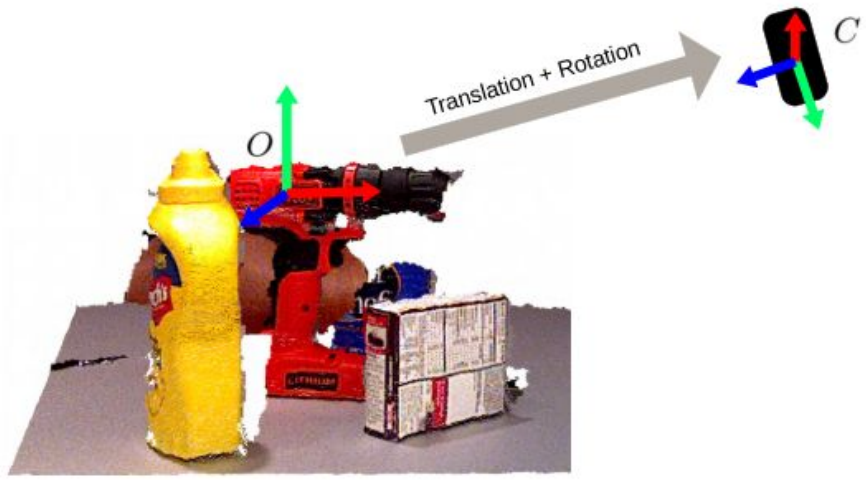


1. Research Objective : Object Tracking for Augmented Reality experience

Object Tracking enables Real-time 360° Augmented Reality experiences around physical objects.

Object Tracking enables AR applications to track and augment several 3D objects simultaneously.

- Once the objects are recognized, users are able to see augmented reality content (3D models, buttons, videos, animations and more) on top of all targets.
- It is also possible to make the digital augmentations interact with each other.
- Object Tracking is the ideal AR feature for enterprise solutions, interactive toys, product enhancement, and more.



6DoF Object Pose Tracking consists of rotation of three degrees of freedom and positional movement of three degrees of freedom.

1. Research Objective : Importance of Depth-aware Object Pose Tracking

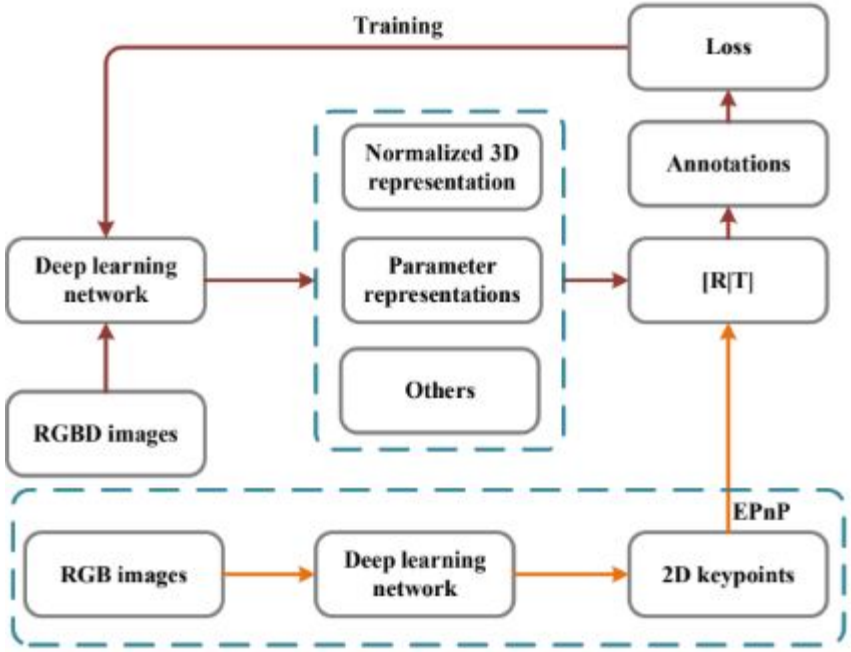
AR glasses use an **RGBD camera** to acquire both color and depth information of the real object and environment.

By comparing the depth value of each pixel corresponding to a real object and a virtual object in an image,

- Based on the **3D spatial relation** and extracting robust features, it is possible to **handle occlusion problems**.
- **Robust to factors such as lighting changes and shooting angles** (causes image blur, reflection, blind spot, cutoff, which may obscure features extracted from images)

To efficiently process RGBD images, an lightweight network and efficient tracking pipeline is needed.

Input RGB Images



2. Contributions of proposed Method

- Efficiently segment the target objects and extract features of the objects through **lightweight 3D keypoints generator** from the **RGBD images which can also handle occlusion problem**

- Compares them with 3D keypoints of existing objects to find out the **6D pose changes(rotation and translation)** and **if the point is the point of the existing object.**

- **Tracks the target objects** and updates the 6D changes between two consecutive frames considering temporal continuity

- **Server-client system** that enables real-time operation on mobile devices such as hololens by using efficient and lightweight network(EfficientNet-Lite) for object detection and segmentation.

→ Handling truncated and occluded scenes

→ Efficient, Real-time Tracking

→ Able to track Category-level Unseen Object + Instance-aware

→ Available for AR application (3D interaction, digital twin management)

→ Low computational cost and Reduce latency

→ Enable Content sharing and Collaboration using AR mobile devices

2. Literature Review : Previous Research and Limitation

SOTA of Open Source SDK for Augmented Reality

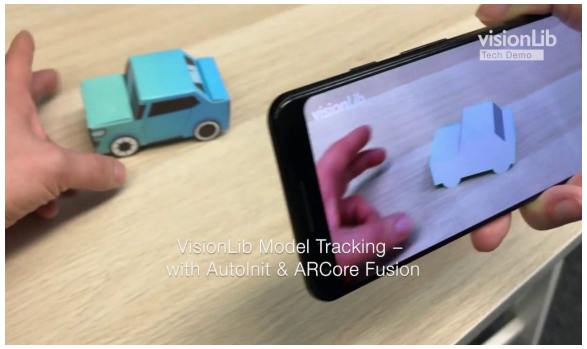
- **Edge-based Approach**

- Traces the silhouette or edge of the object and cannot effectively deal with 3D spatial relations
- Limiting an application to the visibility range of the tracking target and being very sensitive to occlusions of that target

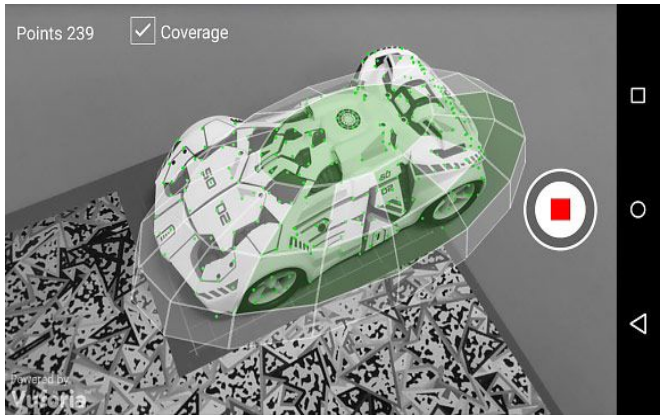
- **Model-based Approach**

- Model-aware methods require 3D CAD models of objects in post-processing
- Hard to get robust features and identify objects
- Use expensive algorithms to estimate pose such as RANSAC, PnP, ICP

VisionLib

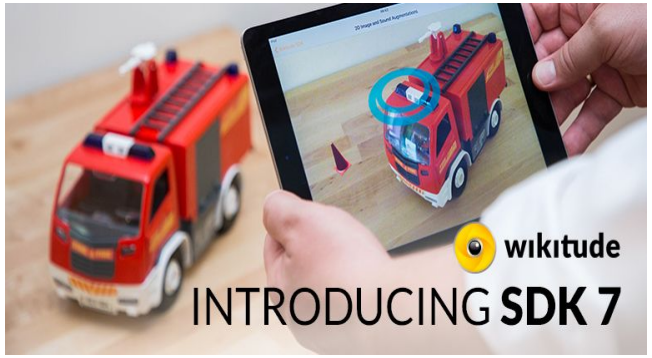


Vuforia Object Anchor



ARKit

Wikitude



2. Literature Review : Previous Research and Limitation

SOTA of Object Tracking Algorithm in Computer Vision

- **Keypoint-based Approach**

RGB-based

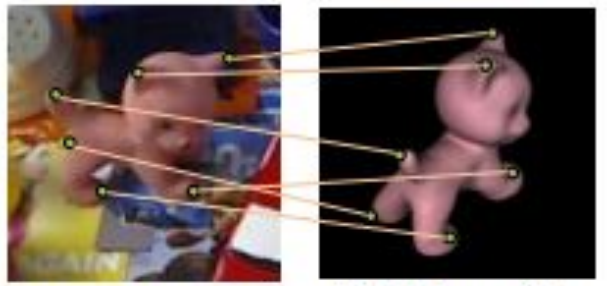
- 9Dof object bounding box is recovered by solving the EPnP problem
- Achieve real-time performance and have the potential of being applied to terminals such as mobile phones.
- Cannot estimate the absolute depth of the object from the camera

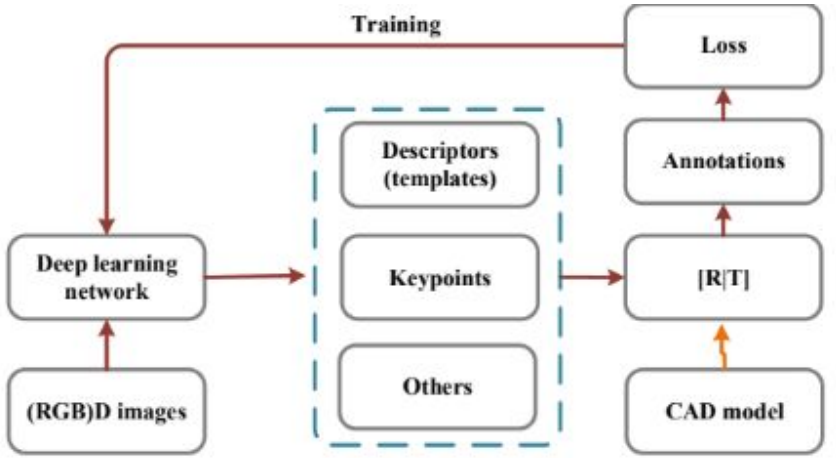
Ex) MobilePose, Objectron(Google)

RGBD-based

- Fully leverages color and depth information
- Robustness in heavy occlusion and changing lighting conditions.
- Need for designing more lightweight network architectures to prompt real-time performance

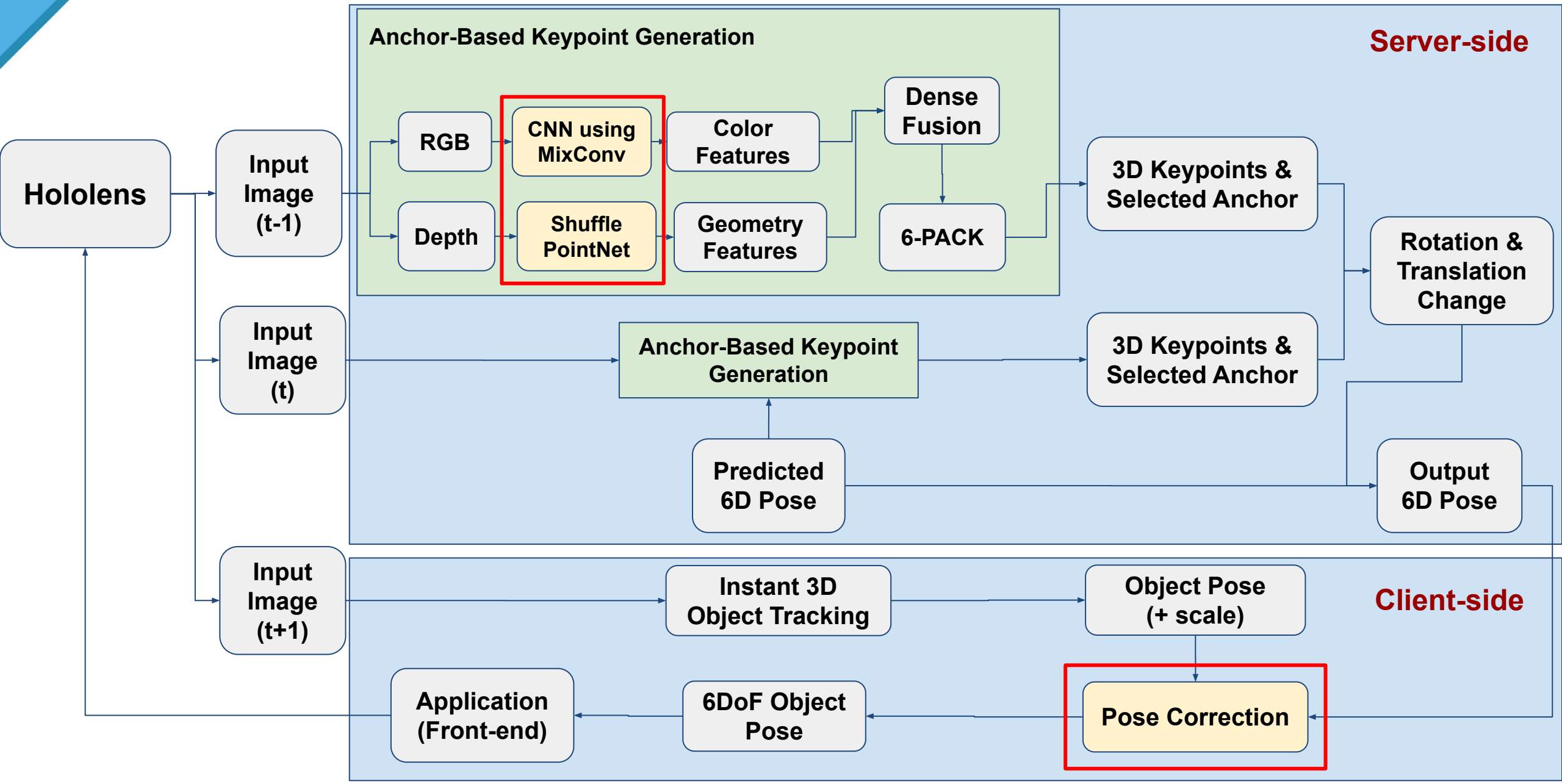
Ex) DenseFusion, 6-PACK, BundleTrack





3. Implementation of Method Proposal

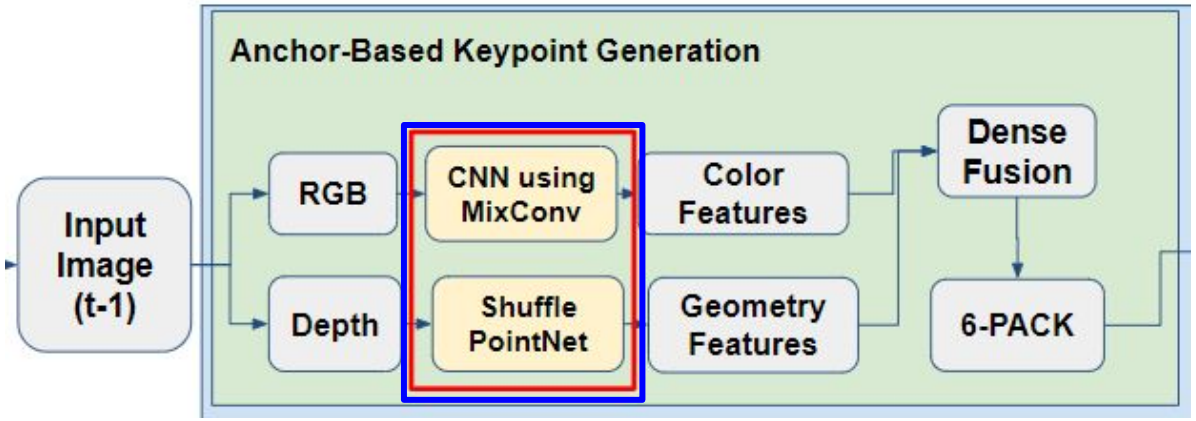
: Main Contribution



3. Implementation of Method - Feature Extraction

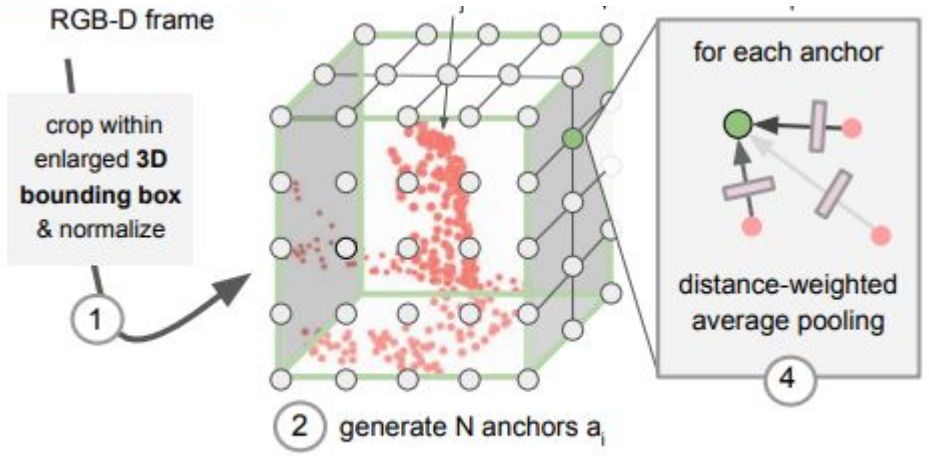
① Color Feature Extraction (CNN using MixConv)

- extract the appearance information from RGB
- use MixConv which has smaller quantity of parameters and better capability of aggregating color features



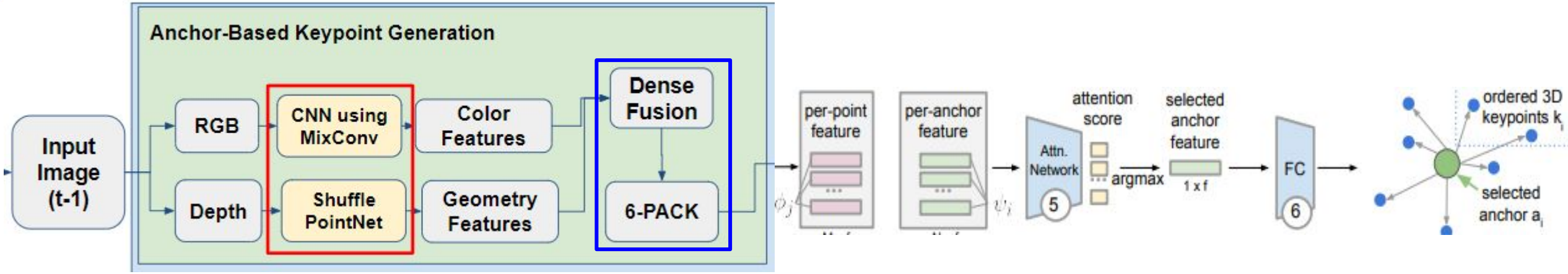
② Geometric Feature Extraction (ShufflePointNet)

- extracts the geometry information in Point clouds
- exploits fine-grained local features
- but also reduce redundancies using group convolution and channel shuffle operation



Anchor-Based Keypoint Generation

3. Implementation of Method - Anchor-Based Keypoint Generation

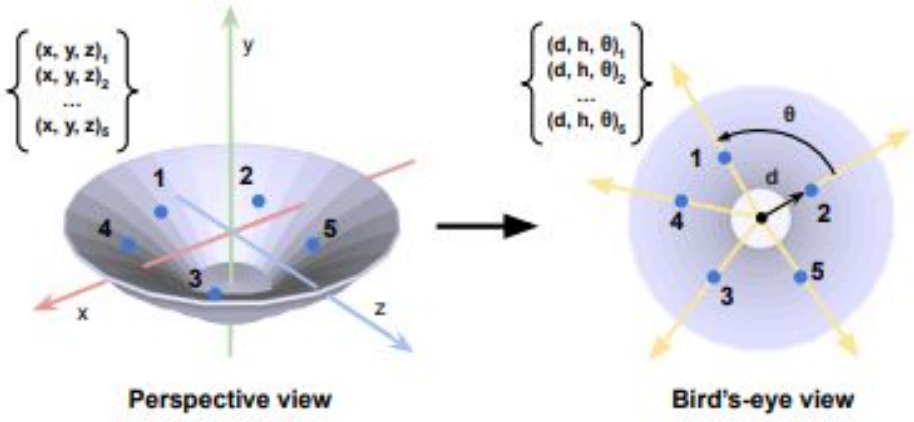


③ Dense Fusion (Pixel-wise Feature Fusion)

- fuse color features and geometry features embeddings
- obtain the combined feature for each point

④ 6-PACK (3D Keypoint Generation)

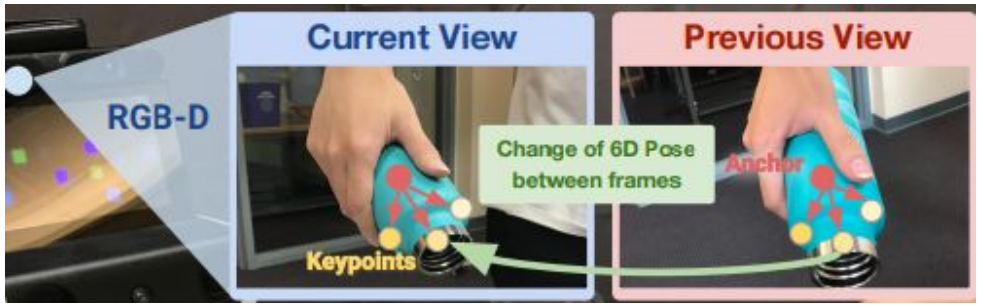
- keypoint generation neural network that uses as input the anchor feature
- generates a K × 3 dimensional output containing an ordered list of keypoints.



3. Implementation of Method

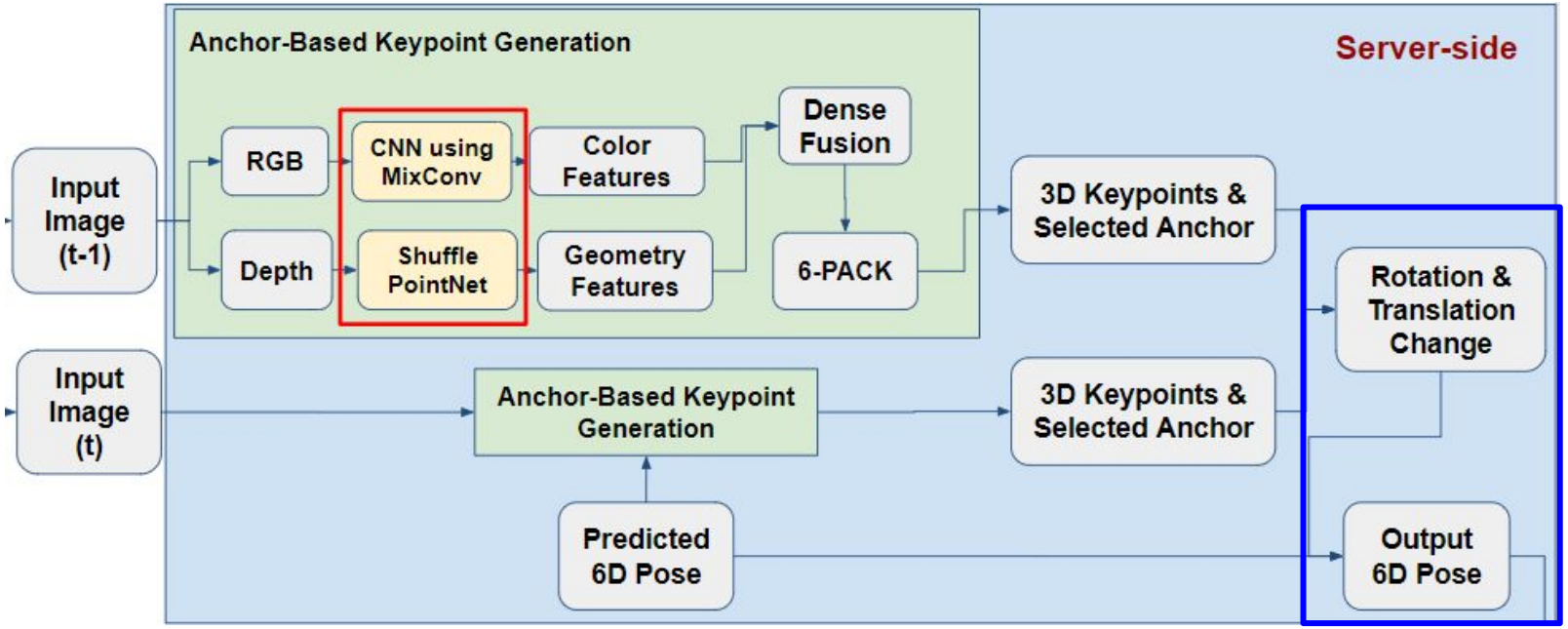
⑤ 6D Pose tracking with inter-frame change in Pose (Server-side)

- Anchor-based keypoint generation is applied on the previous and current frames
- 3D keypoints are generated in each of two consecutive frames
- Obtain two sets of ordered 3D keypoints to compute the inter-frame change in pose.



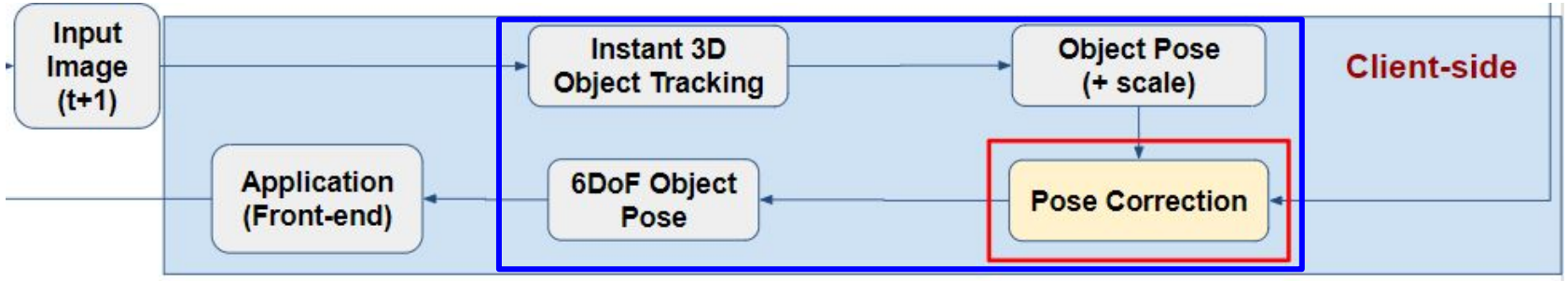
Main Advantages

1) use not only Color but also Geometry feature information
2) Robust to occlusion and light change
3) High accuracy with low computation cost



3. Implementation of Method

Instant 3D Object Tracking & Pose correction (Client-side)



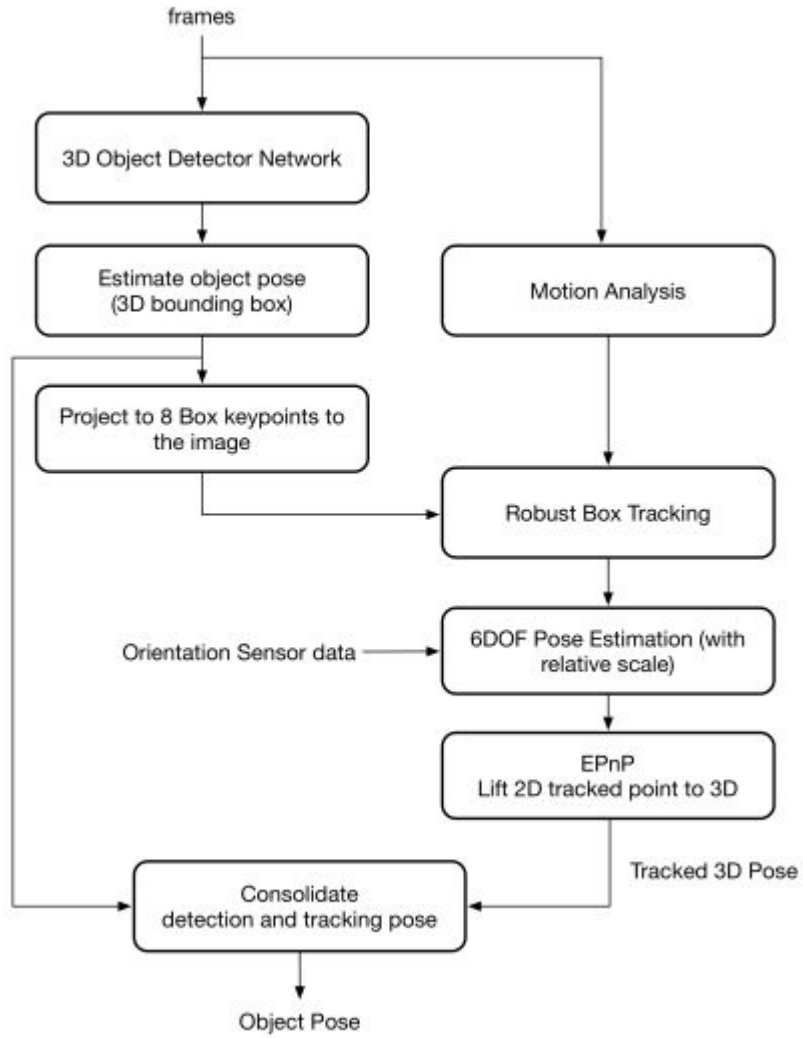
⑥ Instant 3D Object Tracking

- 3D object tracking that enables real-time instant 3D bounding box tracking on mobile devices(Mobilepose)
- Real-time performance
- 9DoF tracking on mobile

⑦ Pose Correction

- compare between the estimated Object Pose from Server and Instant 3D object tracking module
- use FPGM(Filter Pruning via Geometric Median) Algorithm to prune filters with redundancy via geometric median
- refine the poses while improving real-time performance

Instant 3D Object Tracking



4. Technical Evaluation - NOCS-REAL275 Dataset

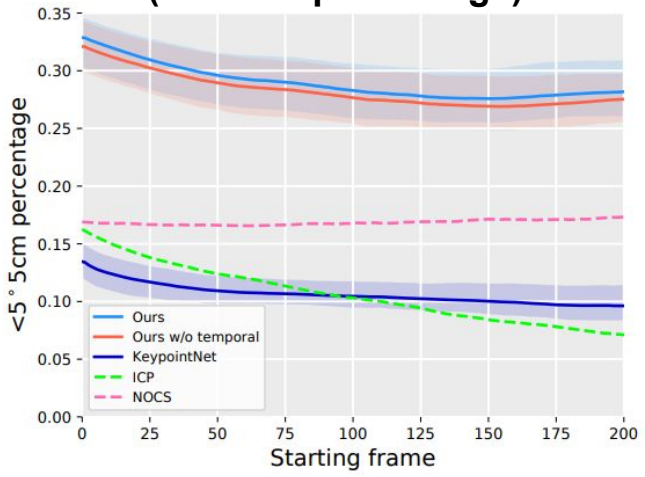
Quantitative Evaluation Metrics of 6D Pose

1) **5°5 cm :** the percentage of tracking results with orientation error < 5° and translation error < 5 cm
2) **3D IoU25 :** percentage of volume overlap between the prediction and ground-truth 3D bounding box that is larger than 25%
3) **Rerr :** mean of the orientation error in degrees
4) **Terr :** mean of the translation error in centimeters

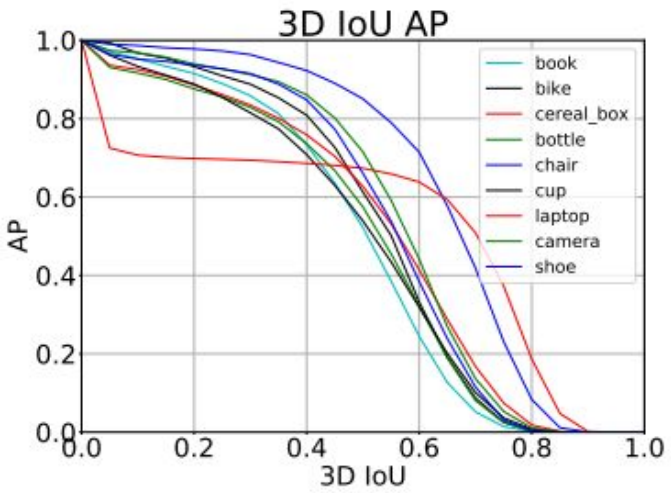
QUANTITATIVE EVALUATION OF 6D POSE ON NOCS-REAL275

		NOCS [46]	ICP [50]	Keypoint Net [41]	Ours w/o temporal	Ours
bottle	5°5cm	5.5	10.1	5.9	23.7	**24.5**
	IoU25	48.7	29.9	23.1	**92.0**	91.1
	R_{err}	25.6	48.0	28.5	15.7	**15.6**
	T_{err}	14.4	15.7	9.5	4.2	**4.0**
bowl	5°5cm	**62.2**	40.3	16.8	53.0	55.0
	IoU25	99.6	79.7	74.7	**100.0**	**100.0**
	R_{err}	**4.7**	19.0	9.8	5.3	5.2
	T_{err}	**1.2**	4.7	8.2	1.6	1.7
camera	5°5cm	0.6	**12.6**	1.8	8.4	10.1
	IoU25	90.6	53.1	30.9	**91.0**	87.6
	R_{err}	**33.8**	80.5	45.2	43.9	35.7
	T_{err}	**3.1**	12.2	8.5	5.5	5.6
can	5°5cm	7.1	17.2	4.3	**25.0**	22.6
	IoU25	77.0	40.5	42.6	89.9	**92.6**
	R_{err}	16.9	47.1	28.8	**12.5**	13.9
	T_{err}	**4.0**	9.4	13.1	5.0	4.8
laptop	5°5cm	25.5	14.8	49.2	62.4	**63.5**
	IoU25	94.7	50.9	94.6	97.8	**98.1**
	R_{err}	8.6	37.7	6.5	4.9	**4.7**
	T_{err}	**2.4**	9.2	4.4	2.5	2.5
mug	5°5cm	0.9	6.2	3.1	22.4	**24.1**
	IoU25	82.8	27.7	52.0	**100.0**	95.2
	R_{err}	31.5	56.3	61.2	**20.3**	21.3
	T_{err}	4.0	9.2	6.7	**1.8**	2.3
Overall	5°5cm	17.0	16.9	13.5	32.5	**33.3**
	IoU25	82.2	47.0	53.0	**95.1**	94.2
	R_{err}	20.2	48.1	30.0	17.1	**16.0**
	T_{err}	4.9	10.5	8.4	**3.4**	3.5

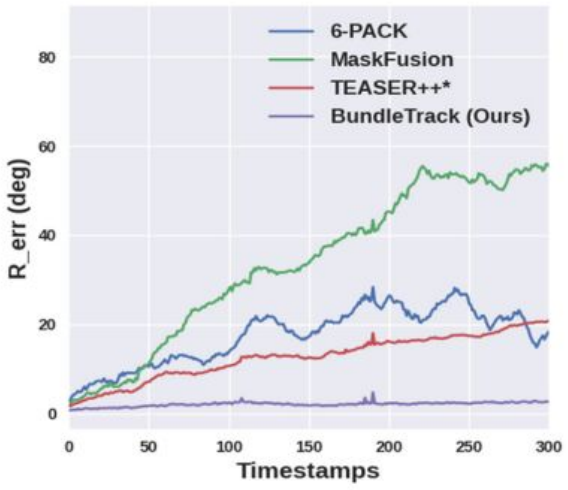
mean success rate (<5°5 cm percentage)



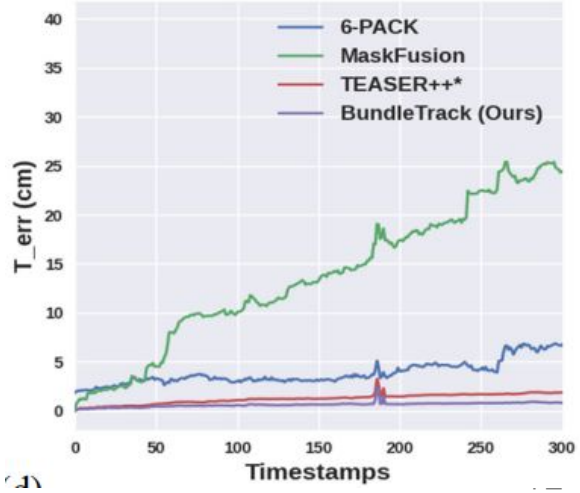
3D IoU25



Rerr



Terr



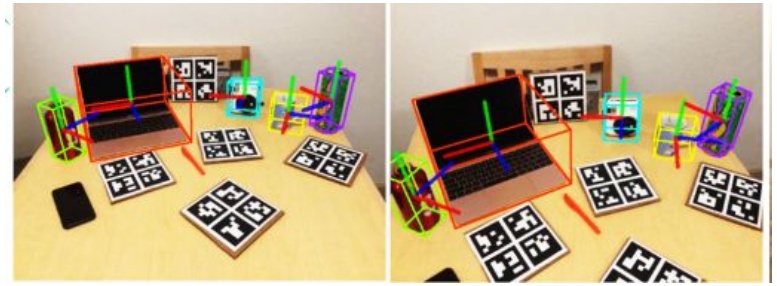
4. Technical Evaluation

Qualitative Comparison between Our Method and Others

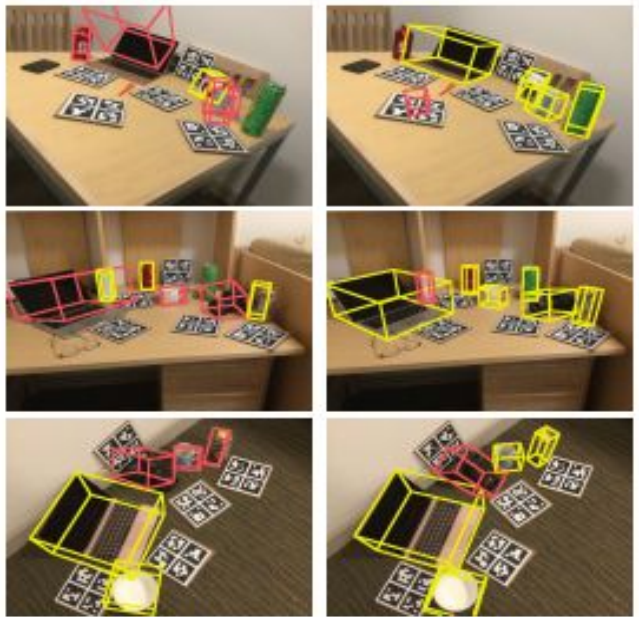
1) 6DoF Object Pose in xyz axes Tracking frames
2) Tracking a 3D bounding box with both camera motion and object movement
3) Visualization of the Keypoints

   : Generated and matched keypoints in two largely separated frames.
3) Fitting and rendering a mesh model to the detected objects for AR Applications

6DoF Object Pose Tracking



3D Bounding Box



Keypoints Visualization

Rendering Mesh Model

5. User Study

Test : 100 participants (50 male, 50 female) with a normal or corrected-to-normal vision watch and evaluate 3D Bounding Box Tracking Videos and Augmented Reality application

Hardware : Microsoft Hololens 2

Model Comparison : 6-PACK, BundleTrack, Our model

Object categories : 6 objects including bottle, bowl, camera, can, laptop, and mug

Video Contents

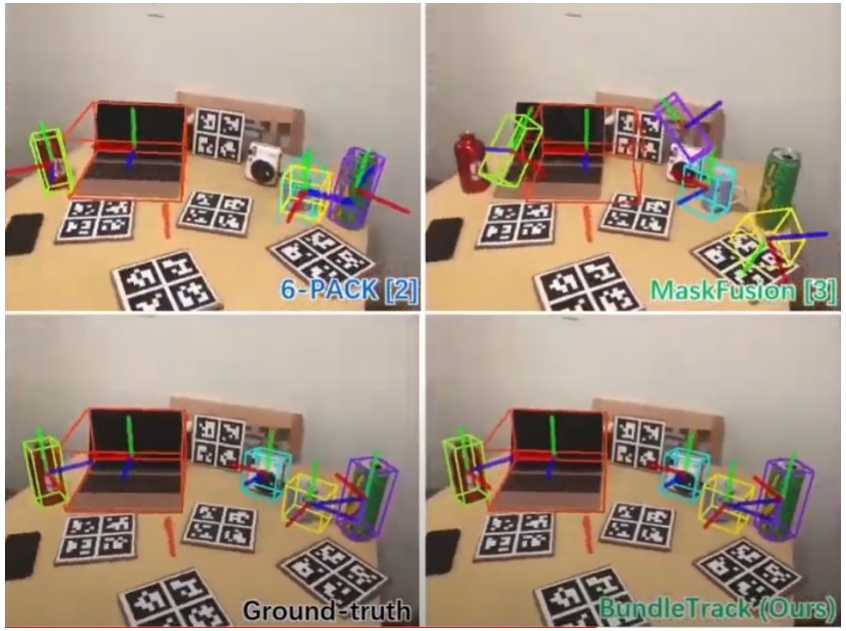
1) **6DoF Pose and 3D Bounding Box Tracking Real videos** depicting in total two different (unseen) instances for each object category
   The total number of videos = 2(unseen instances)*6(object categories)*3(models) = 36

2) **Augmented Reality application** that renders mesh models overlaying to a real object for each object category
   The total number of objects in one scene = 6

5. User Study

6DoF Pose and 3D Bounding Box Tracking Real videos

Questionnaire:



① Perceptual Realism

Which of the object tracking model is the best?

② Smoothness of Object tracking

Do you think the provided object tracking video looks smoothly track the target object?

Likert scale : ranging from 1 (low naturalness) to 5 (high naturalness)

③ Give Scores to the Videos of each Models

6DoF Pose and 3D Bounding Box Tracking Real videos

The total number of videos to evaluate = 2(unseen instances)*6(object categories)*3(models) = 36

Score Scale : ranging from

5. User Study

Augmented Reality Application Contents

Questionnaire:

① Perceptual Realism

Which of the AR Contents depicting objects with 3D rendered mesh is the best?

② Naturalness of AR Contents

Do you think the provided AR contents looks naturally overlayed to target object?

Likert scale : ranging from 1 (low naturalness) to 5 (high naturalness)

③ User Satisfaction

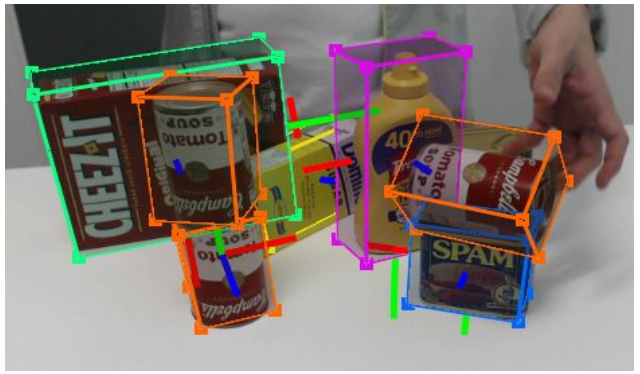
Did you satisfy while using AR application?

Measure : ranging from 1(Not Satisfied) to 5 (Very Satisfied)

6. Expected Results

Depth aware Real-time 6DoF Object tracking

for Augmented Reality



Contribution

- Achieve Lightweight, Stable and Fast enough Object 6DoF Pose estimation and Tracking for Real-time operation on Mobile devices like Hololens
- Color and Geometry embedding for Feature Generation and 3D Keypoints matching that can use geometry information by not only using RGB but also Depth data
- Improve Speed performance of Pose tracking by computing the inter-frame change in 6D Pose (Rotation and Translation)
- Solve Occlusion and Light-changing problem
- Improve Accuracy with Dual-operating Pose tracking pipeline
- Enable Real-time Interaction between real object and AR contents wearing AR HMD and gives Fast Feedbacks

7. Potential Application

Object pose detection and tracking has recently attracted increasing attention due to its wide applications in many areas, such as autonomous driving, robotics, and augmented reality.

Objects that can be tracked as AR target Objects include but are not limited to :

- **Education**
  It can be used for various visual learning materials such as visualization or simulation of information on learning subjects in schools.
- **Tourism**
  As an entertainment content of a museum, information can be visualized in historical relics or interactive content can be developed.
- **Industrial objects**
  It can be used to classify industrial objects or obtain information about the environment in factories, and can be used by workers to quickly learn the working environment.

7. Potential Application

Object pose detection and tracking has recently attracted increasing attention due to its wide applications in many areas, such as autonomous driving, robotics, and augmented reality.

Objects that can be tracked as AR target Objects include but are not limited to :

- **Interactive Toys**

  Build engaging games by combining toys with the power of augmented reality. Object tracking allows you to track 3D targets and display virtual creatures, game guidelines, buttons and other interactive content on physical toys. By adding 3D animations, and videos, your product or app can take the players' engagement to a whole new level.

- **Remote Assistance**

  Brands and enterprises are turning AR to remotely support consumers and workers when an expert is not around. Workers can scan multiple objects to check for faulty settings and shoppers can get accurate step-by-step guides on how to set up their new appliances

Thank You