

Automatic Image Collection of Objects with Similar Function by Learning Human Grasping Forms

Shinya Morioka, Tadashi Matsuo, Yasuhiro Hiramoto,
Nobutaka Shimada, Yoshiaki Shirai

Ritsumeikan University

Outline

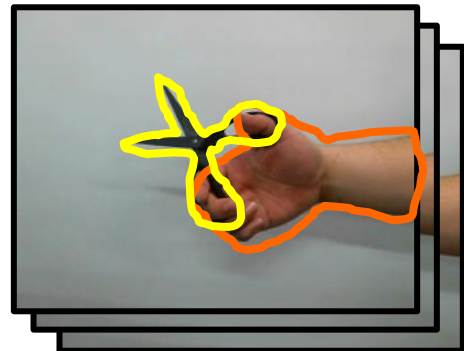
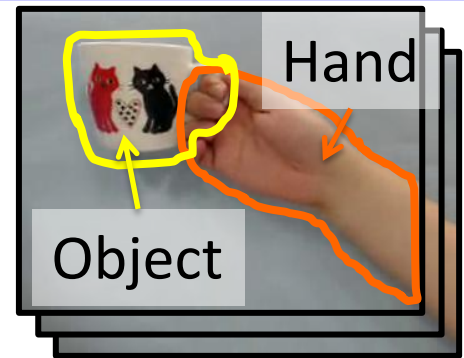
- Motivation
- Related Work
- Proposed Method
- Results
- Conclusions

Motivation(1/2)

- How to classify unknown objects into categories such as “for drinking”, “for cutting” ... ?



To learn this relation, many labeled images are required.



but...

for drinking



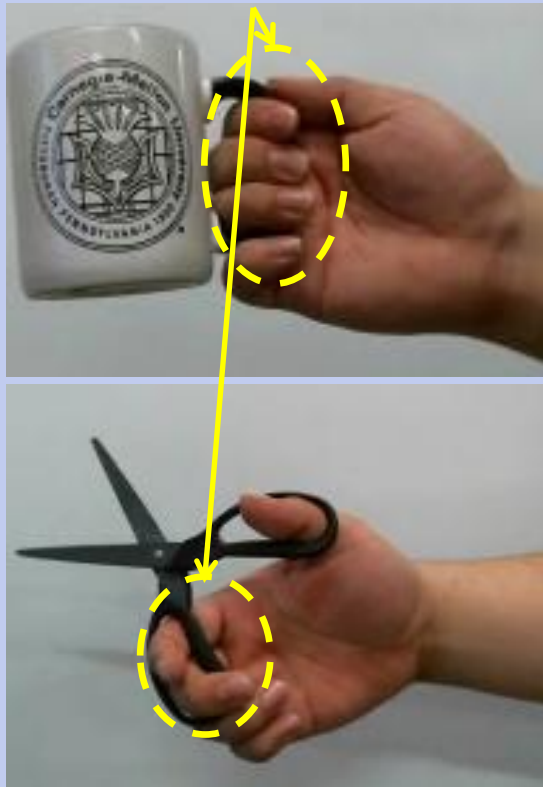
for cutting



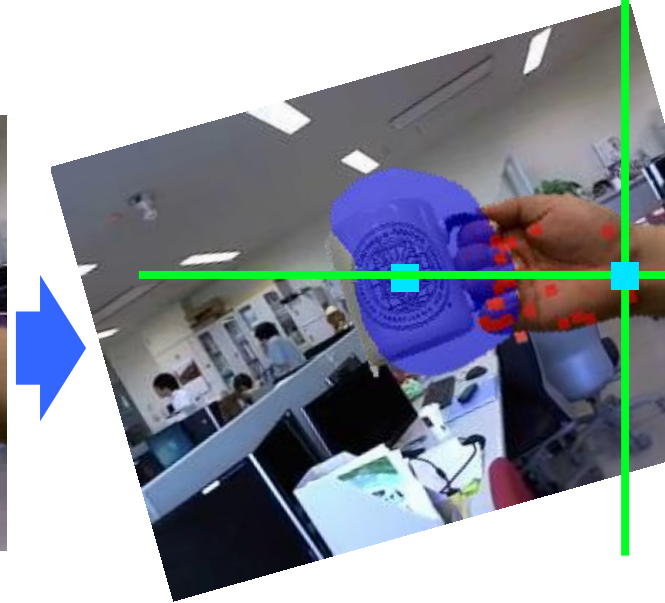
Motivation(2/2)

Object features
may be invisible !

How to collect labeled images?



Extracted
SURF features



Labeled image
normalized with wrist

We propose a method to estimate an object region and standard coordinate system based on the wrist.

Related work

R. Filipovych, E. Ribeiro,

“Recognizing Primitive Interactions by Exploring Actor–Object States”, Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp.1–8(2008).

- They focus on estimation of a time sequence of states when one interact with an object.
- It is based on the relation between local appearances and interaction state.
- It is difficult to estimate an object region in a single image.

We propose a method to estimate an object region and standard coordinate system based on the wrist.

Proposed Method

Training

1. Training a local estimator of wrist position
2. Estimating wrist position
3. Training a local estimator of object position
4. Generating wrist-object coordinate system
5. Training a estimator of object region on the wrist-object coordinate system
6. Estimating an object region

Estimating

2.

4.

6.

Proposed Method(1/6)

Training

1. Training a local estimator of wrist position
2. Estimating wrist position
3. Training a local estimator of object position
4. Generating wrist-object coordinate system
5. Training a estimator of object region on the wrist-object coordinate system
6. Estimating an object region

Estimating

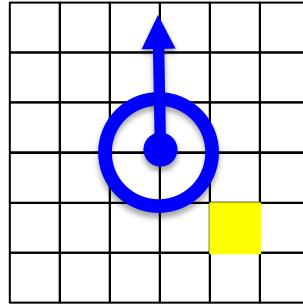
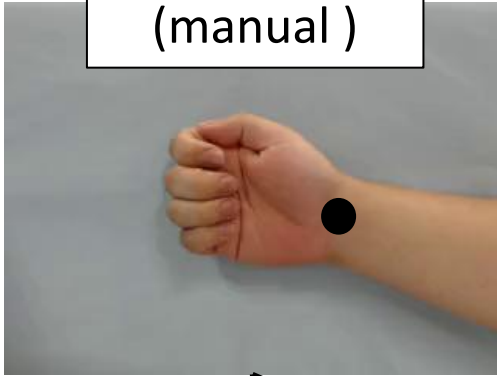
2.

4.

6.

Training a local estimator of wrist position

wrist position
(manual)



Train Randomized Trees(RTs)
with relative wrist positions
for SURF features.

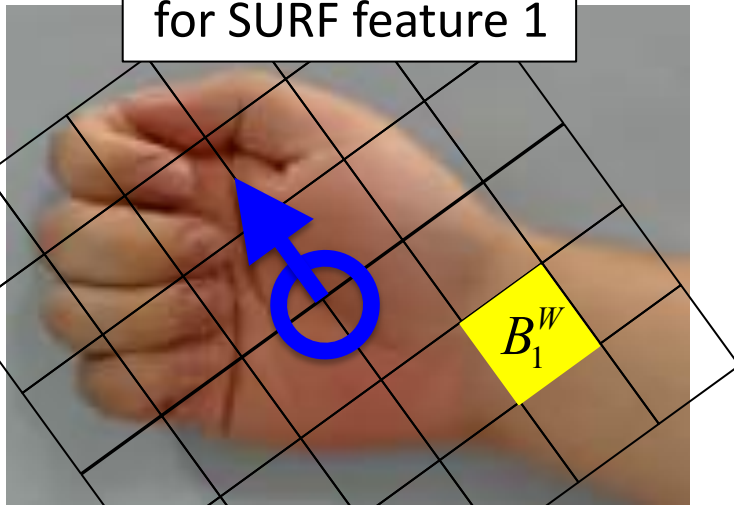


With trained RTs, we can calculate ...

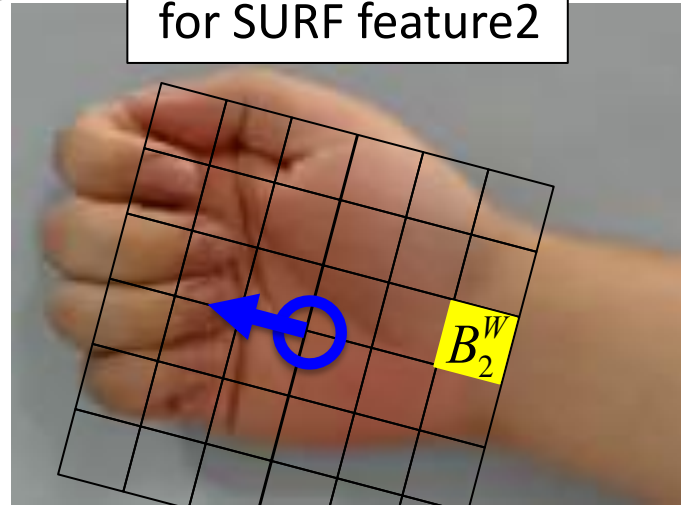
$$\text{Prob}(B = B^W \mid \text{SURF descriptor}(128D))$$

B : block B^W : the block including wrist position

Local coordinate
for SURF feature 1



Local coordinate
for SURF feature2



...

Proposed Method(2/6)

Training

1. Training a local estimator of wrist position
2. Estimating wrist position
3. Training a local estimator of object position
4. Generating wrist-object coordinate system
5. Training a estimator of object region on the wrist-object coordinate system
6. Estimating an object region

Estimating

2.

4.

6.

Estimating wrist position(1/2)

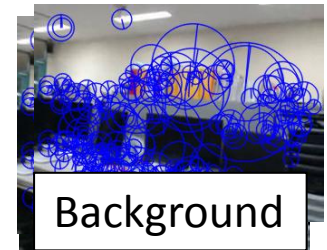
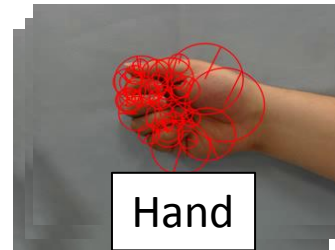
- As preprocess, we remove features apparently from outside of a hand because they interfere the following voting process.

2-Class SVM

Input : SURF descriptor

Output : Hand or Background

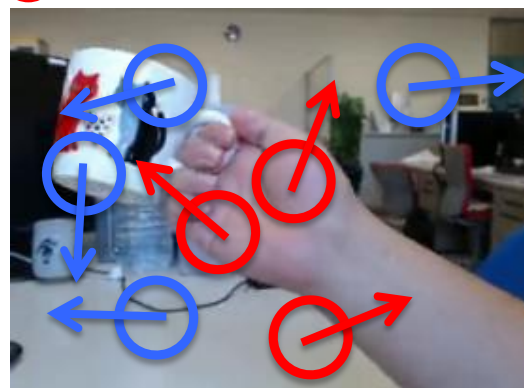
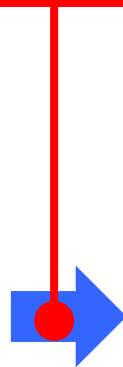
Teachers



○ Hand ○ Background



SURF feature
extraction

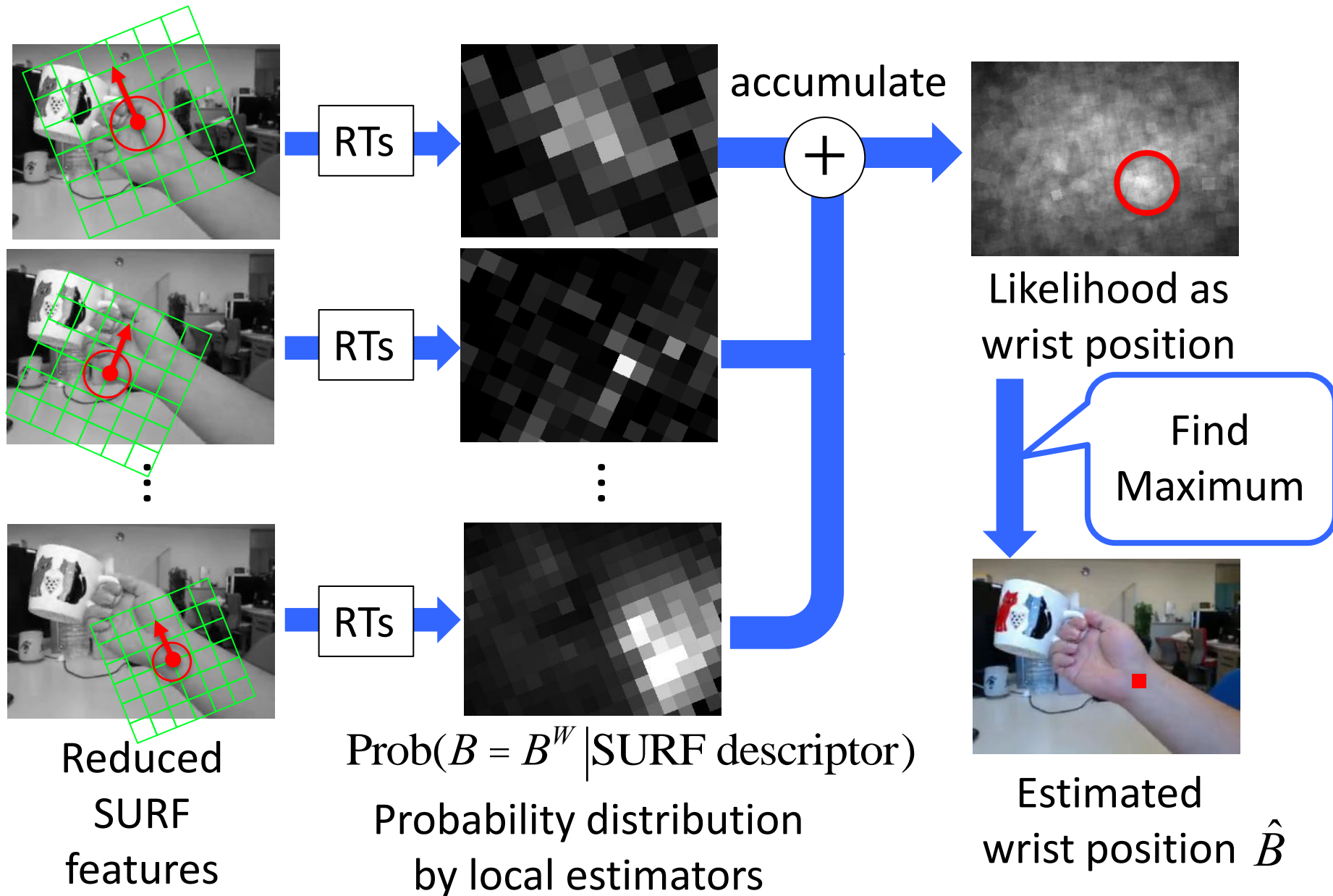


Classification of
features by SVM



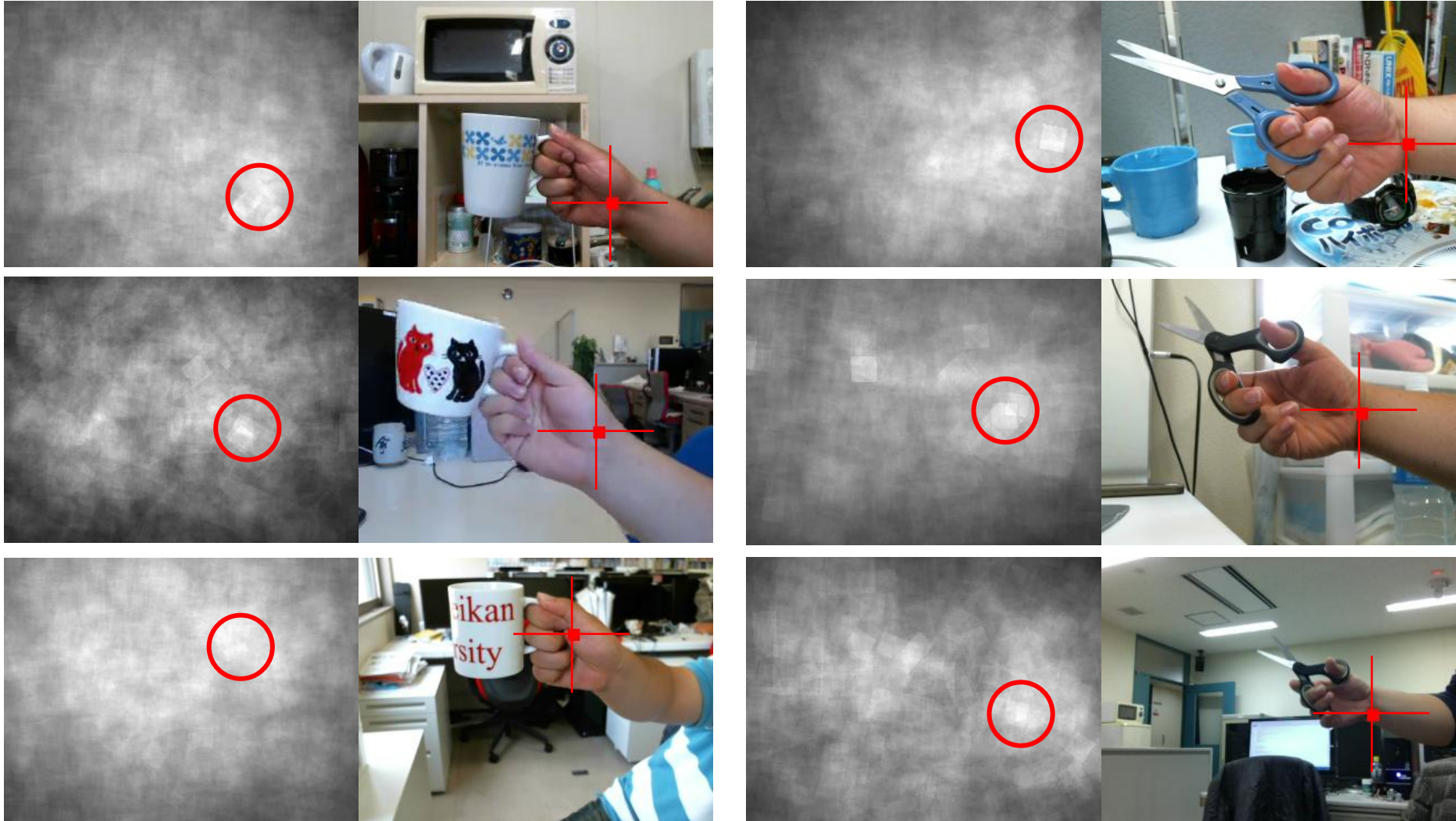
Reduced features
classified as hand

Estimating wrist position(2/2)



RESULTS : estimated wrist position

 : wrist position



Proposed Method(3/6)

Training

1. Training a local estimator of wrist position
2. Estimating wrist position
3. Training a local estimator of object position
4. Generating wrist-object coordinate system
5. Training a estimator of object region on the wrist-object coordinate system
6. Estimating an object region

Estimating

2.

4.

6.

Likelihood as a hand part

Likelihood as a hand part is useful to estimate a hand region and an object region.

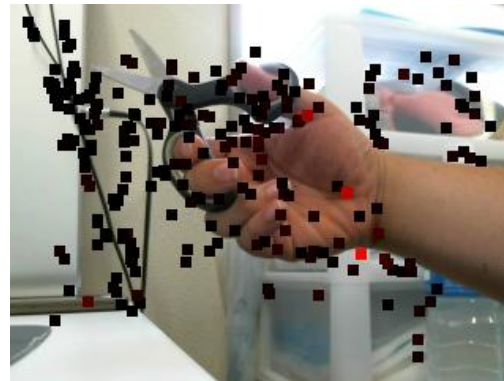
If the j -th SURF feature originates from a hand, $\text{Prob}(B = B^W | j\text{-th SURF descriptor})$ should be high around the estimated wrist position \hat{B} .



We take $h_j \propto \text{Prob}(\hat{B} = B^W | j\text{-th SURF descriptor})$ as a likelihood as a hand part of the j -th SURF feature.

Likelihood h_j as a hand part

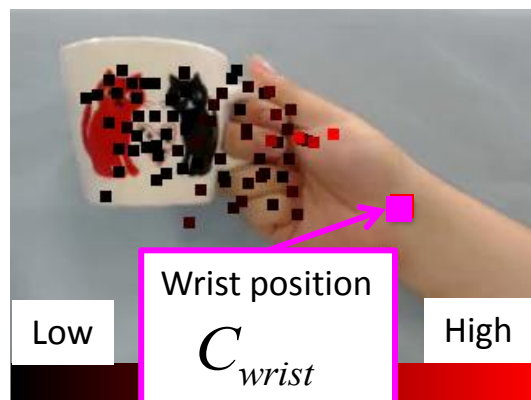
RESULTS : Likelihood as a hand part



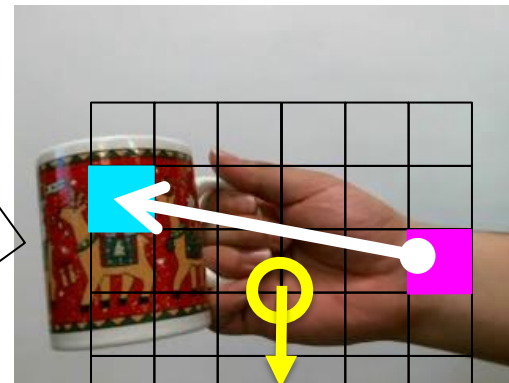
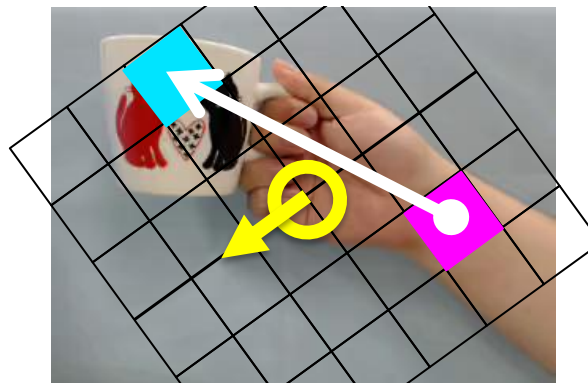
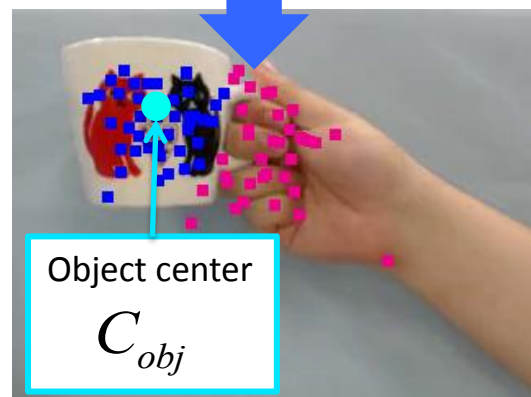
High

Low

Training a local estimator of object center



K-means
clustering
 $(x_j, y_j, h_j), K = 2$



We train RTs with $C_{obj} - C_{wrist}$ represented by each local coordinate.
(Here, we collect the vectors from images with simple background.)



With trained RTs, we can calculate ...

$$\text{Prob}(C_{obj} - C_{wrist} | j\text{-th SURF descriptor})$$

Proposed Method(4/6)

Training

1. Training a local estimator of wrist position
2. Estimating wrist position
3. Training a local estimator of object position
4. Generating wrist-object coordinate system
5. Training a estimator of object region on the wrist-object coordinate system
6. Estimating an object region

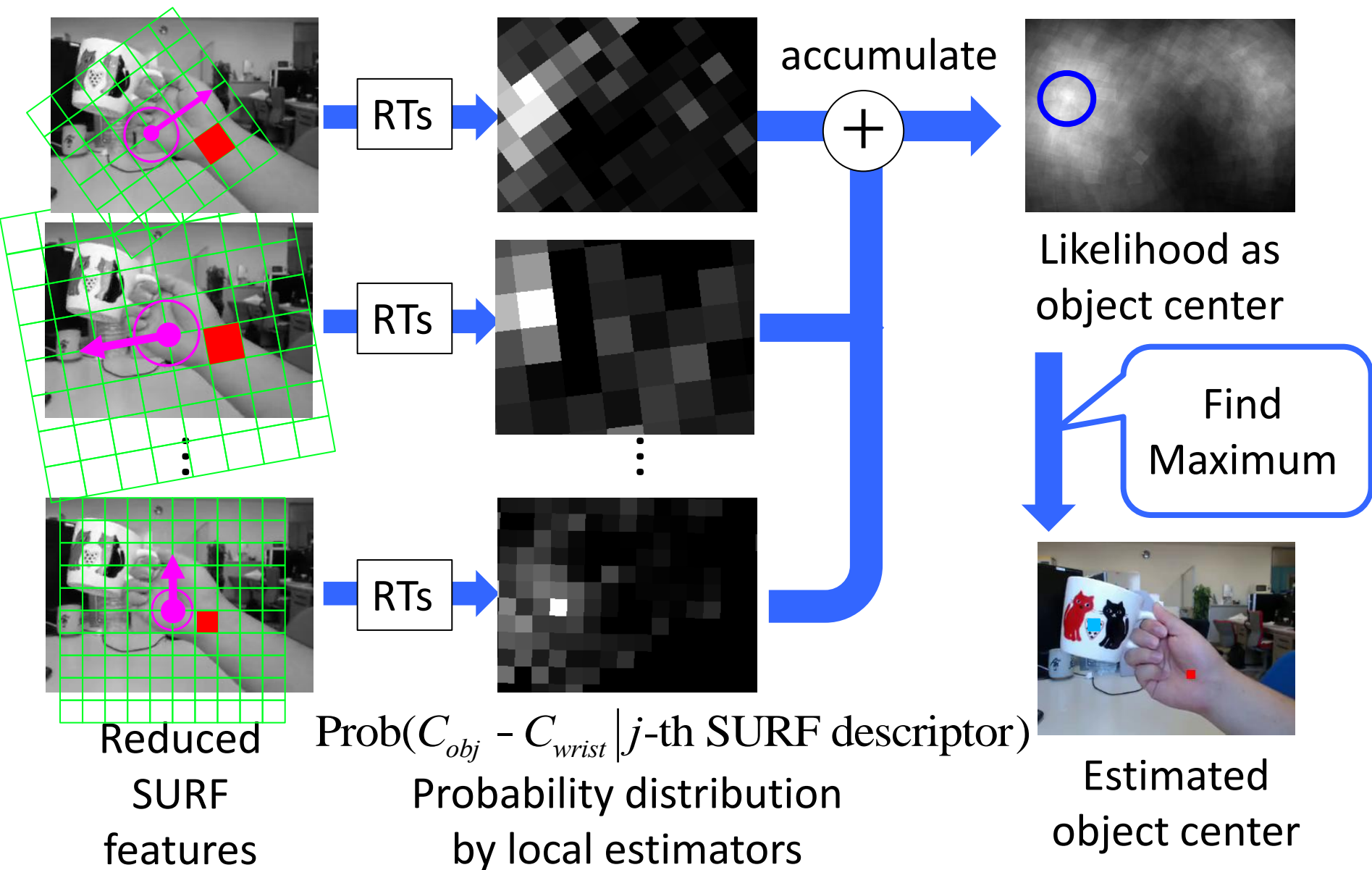
Estimating

2.

4.

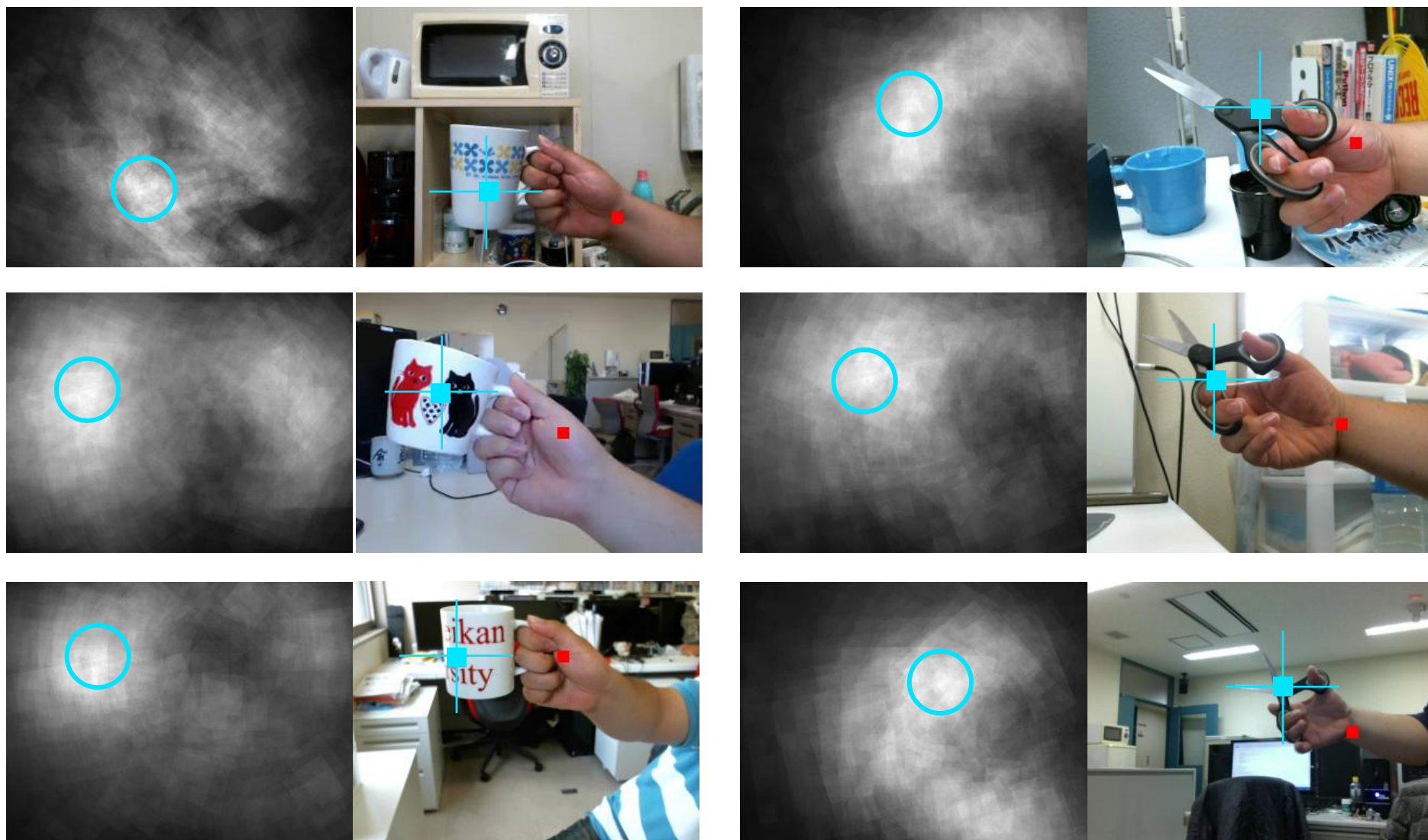
6.

Estimating object center



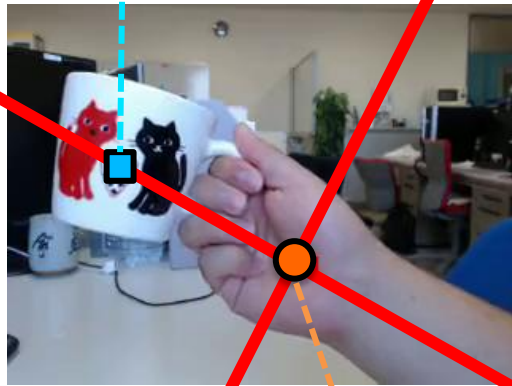
RESULTS : estimated object center

■ : Object center ■ : wrist position



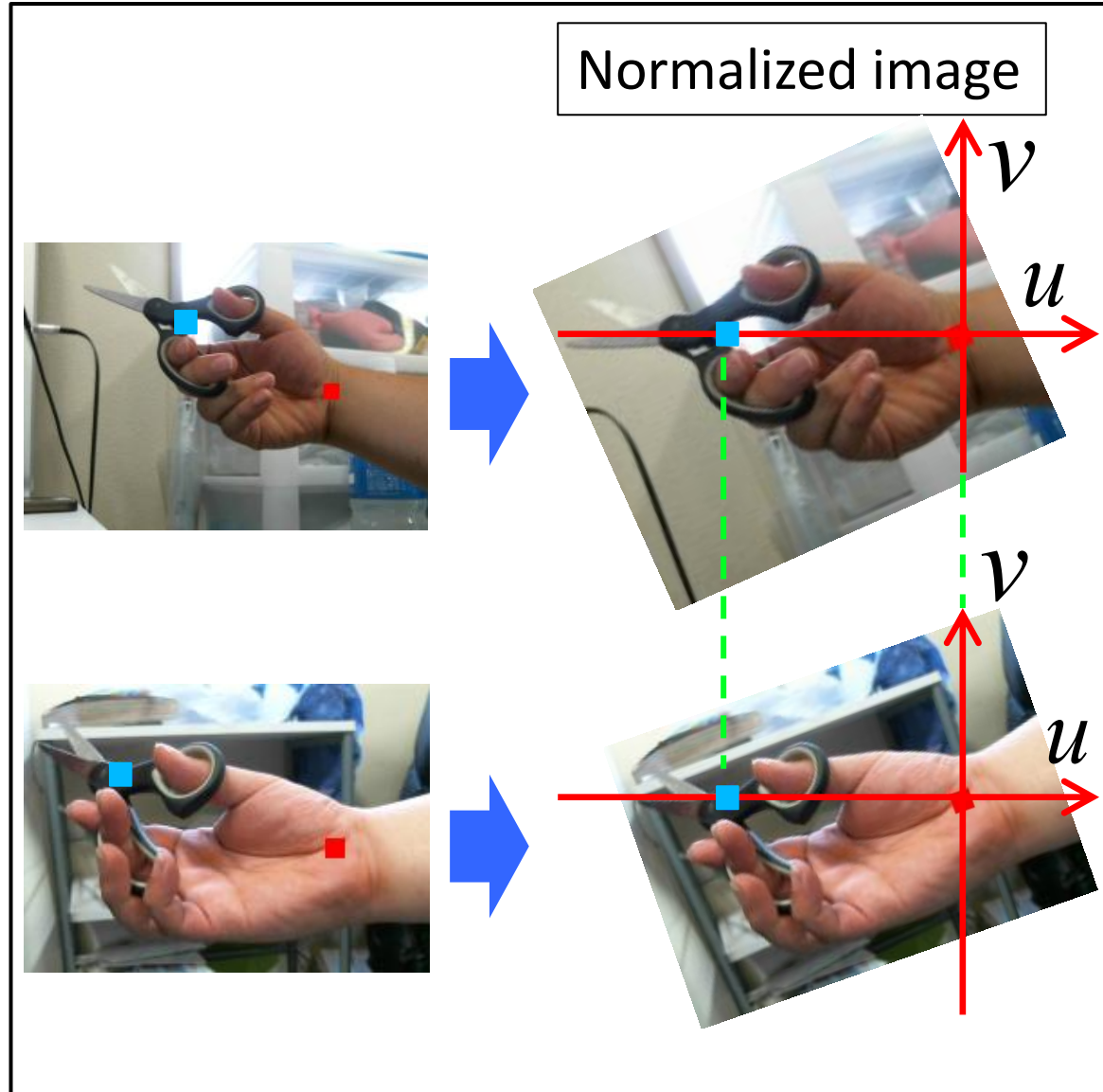
Generating wrist-object coordinate system

object center
 $(u, v) = (-1, 0)$



wrist position
 $(u, v) = (0, 0)$

wrist-object
coordinate system



Proposed Method(5/6)

Training

1. Training a local estimator of wrist position
2. Estimating wrist position
3. Training a local estimator of object position
4. Generating wrist-object coordinate system
5. Training a estimator of object region on the wrist-object coordinate system
6. Estimating an object region

Estimating

2.

4.

6.

Training a estimator of object region on the wrist-object coordinate system

- One-Class SVMs are built for hand and object.

SVM for Hand region

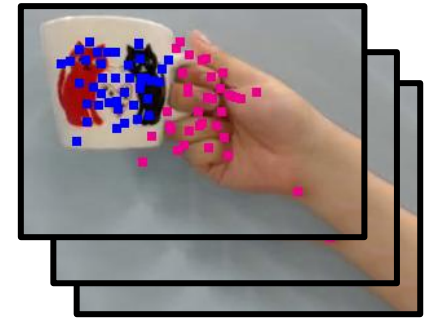
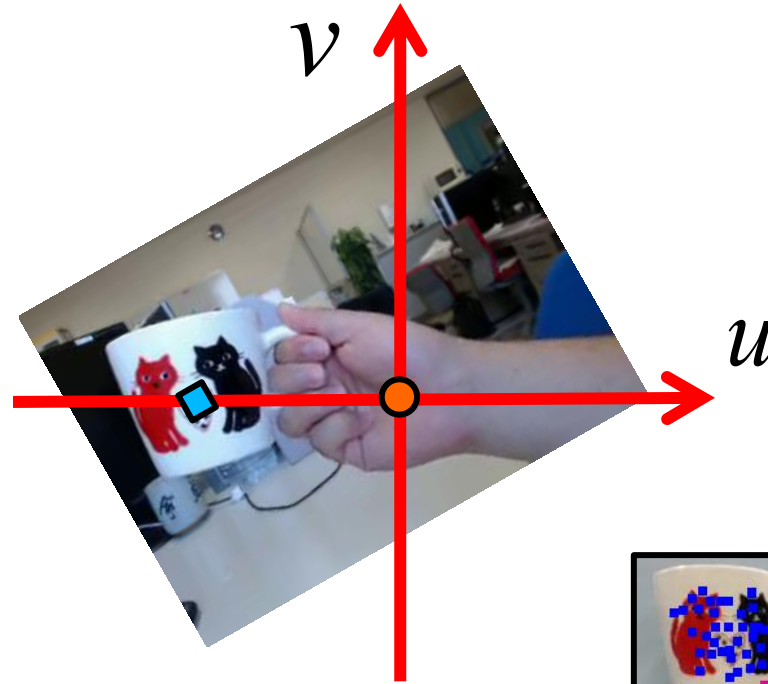
Input : (u_j, v_j, h_j)

Output : Whether (u_j, v_j) is on a hand or not

SVM for Object region

Input : (u, v)

Output : Whether (u, v) is on an object or not



We collect teachers from the K-means clustering result of images with simple background

Classification into hand, object, and background classes

Proposed Method(6/6)

Training

1. Training a local estimator of wrist position
2. Estimating wrist position
3. Training a local estimator of object position
4. Generating wrist-object coordinate system
5. Training a estimator of object region on the wrist-object coordinate system
6. Estimating an object region

Estimating

2.

4.

6.

TOTAL RESULTS (movie)



Hand

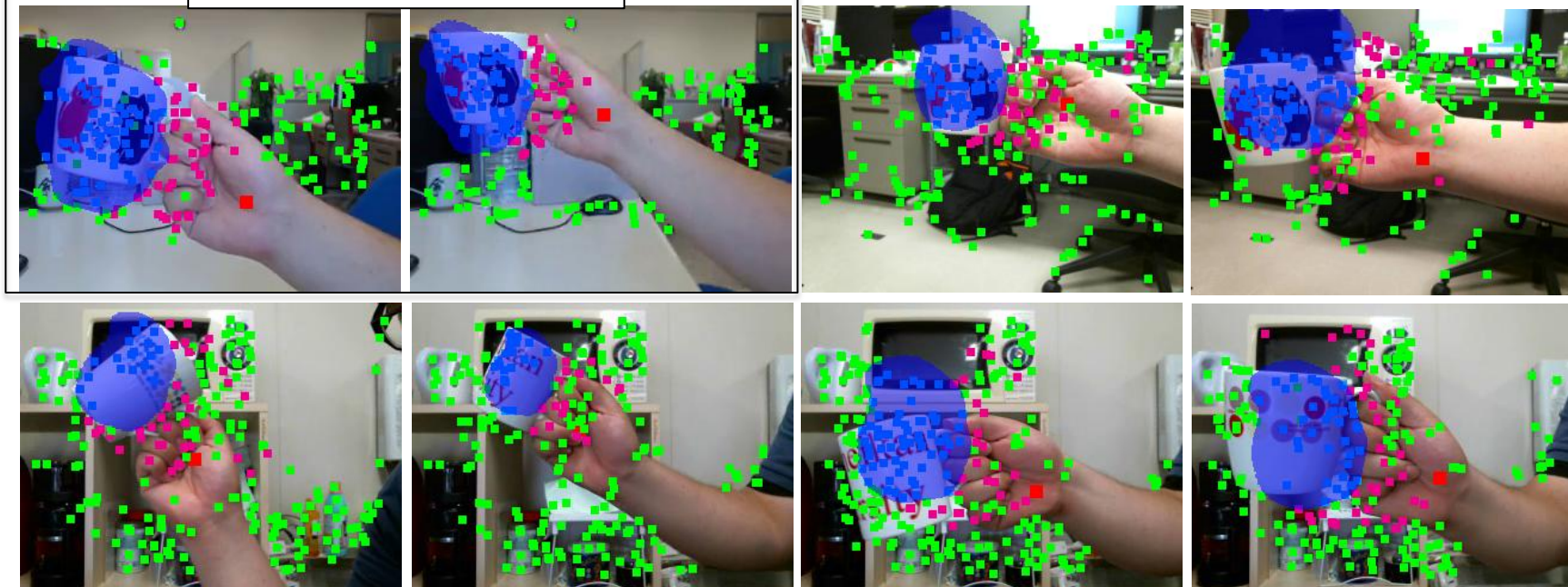


Object

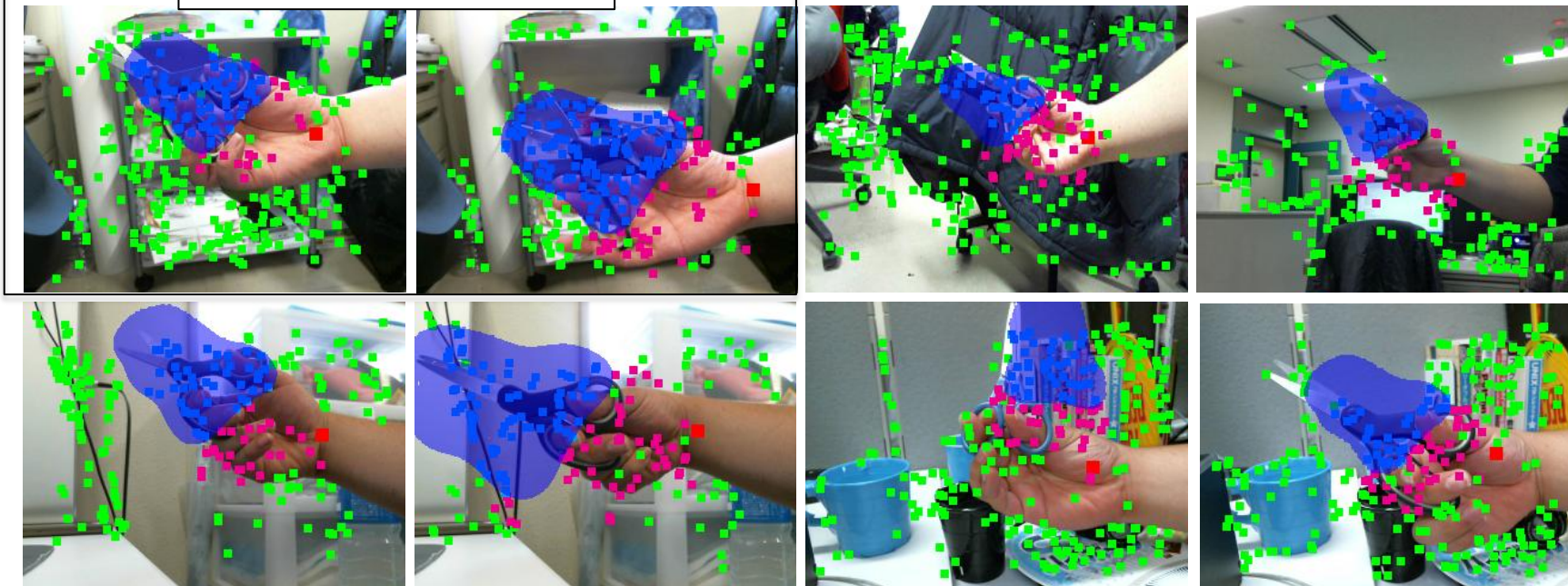


Background

A cup used on training



A cup used on training



Conclusion

- With the proposed method, a wrist can be found and an object center can be estimated from a set of the wrist and local features.
 - The wrist and the object center make a wrist-object coordinate suitable for learning a shape of a grasped object.
-

Future work

- Object recognition by learning the relation between an object and a posture of a hand grasping it.