

Improving a Physical Search System that Detects Even Unknown Displaced Objects Using Image Differences

Kajihara Shin*, Masato Okazaki*, Chika Oshima**, Koichi Nakayama**

*Graduate School of Engineering, Saga University, Japan

**Faculty of Science and Engineering, Saga University, Japan

E-Mail: 20634002@edu.cc.saga-u.ac.jp

knakayama@is.saga-u.ac.jp



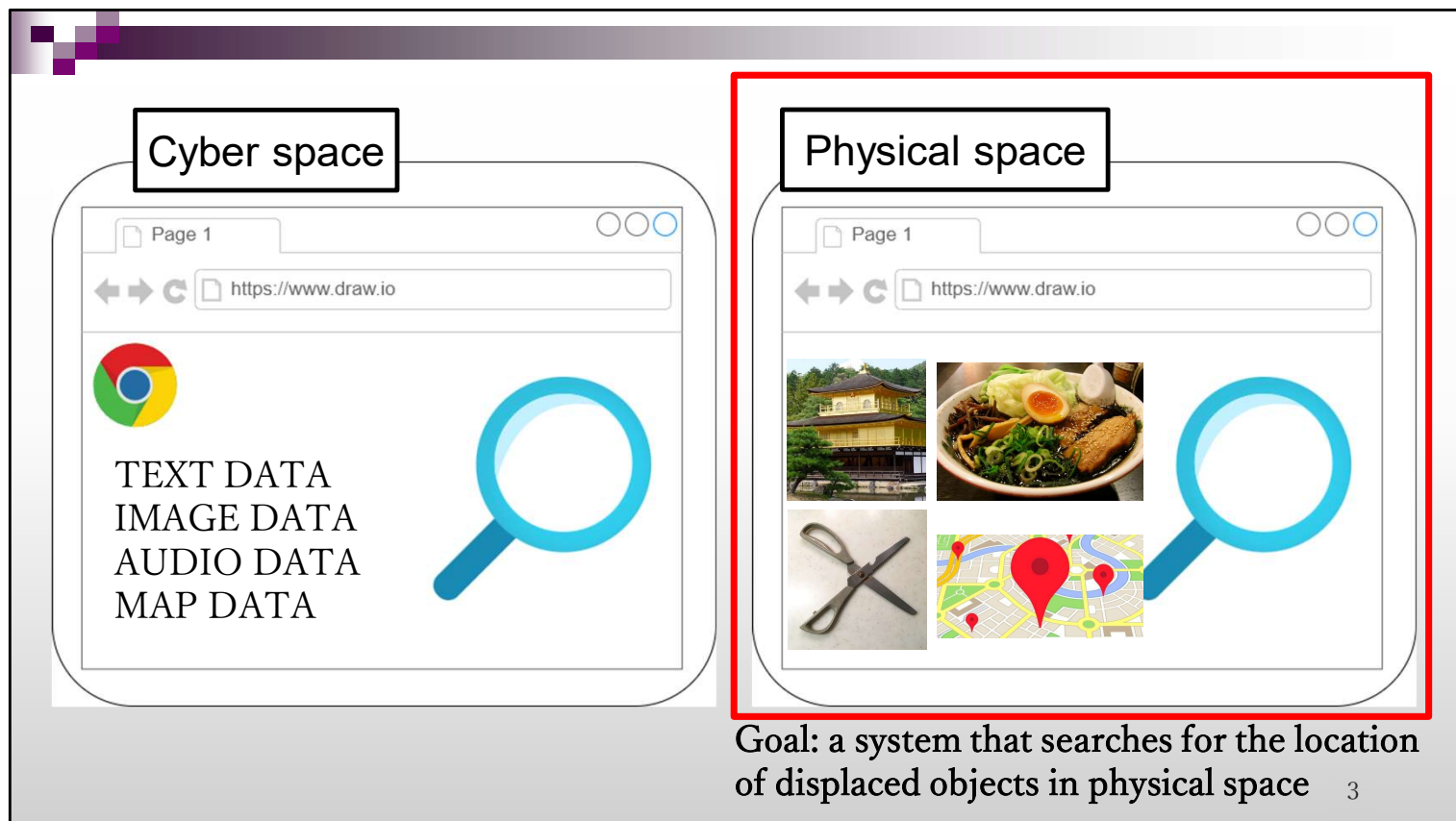
KAJIHARA SHIN

Profile

2017 : Established Locamo-AI LLC.
2018 : Bachelor of Science, Saga University
2020 : Master of Science, Saga University
2020 : Established NEXS Co., Ltd.

In NEXS Co., Ltd.
Doctor of Engineering Candidate, March 2023

Nationality : Japan
Area of expertise : Web development

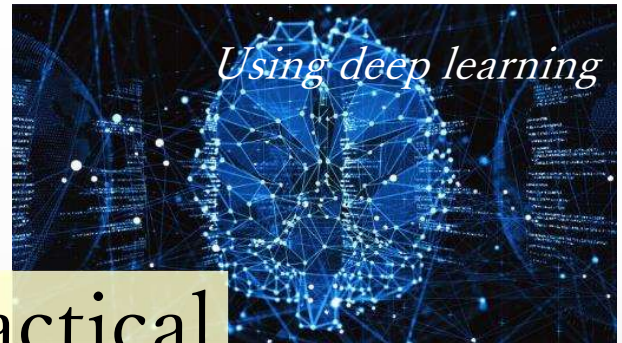


The spread of search engines, such as Google, on the Internet made it easy to indicate where a web page containing a certain keyword is located (i.e., URL) in cyberspace. Currently, cyber-physical system (CPS) has become a hot topic. CPS collects various data in the real world (physical space) with sensor networks. The data are analyzed in cyberspace, and the information created there. The information is used to revitalize industry and solve social problems. However, it is still not easy to search for specific items in the physical space we live in. We are aiming at a system that can search the location of displaced objects in physical space.

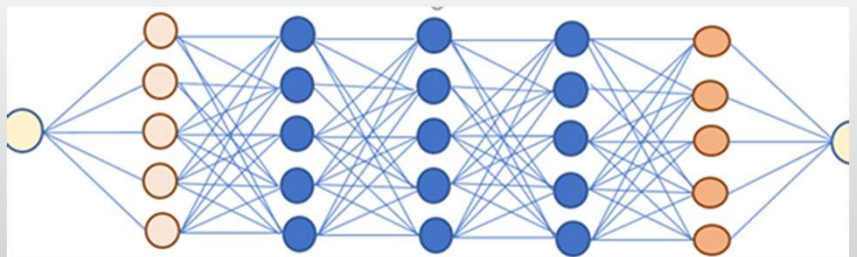
Methods that identify the location of the objects



Electronic tag



Unpractical



4

An electronic tag is one of methods that identify the location of the objects. However, it is not practical to attach such devices to all objects in a room. Recognizing objects using deep learning is also not very useful, because learning all objects is not practical.

PSS detects displaced objects in a physical space using only two cameras

- An analogy to “google search” in cyberspace
- Record various displaced objects based on images difference detection technology
- Label displaced objects with “feature clustering numbers,” color, displaced time.
- Show where the object last displaced as a search result

S. Kajihara, et al. ``Proposal and verification of a physical search system that does not require pre-learning data and sensors other than cameras," IPSJ Transactions on digital practices, 2022. (in Japanese)

5

Therefore, we have developed a Physical Search System (PSS) that uses only two cameras as a sensor and can find unknown objects without additional learning. This system is conceived as an analogy to google search in cyberspace. PSS has the following characteristics,

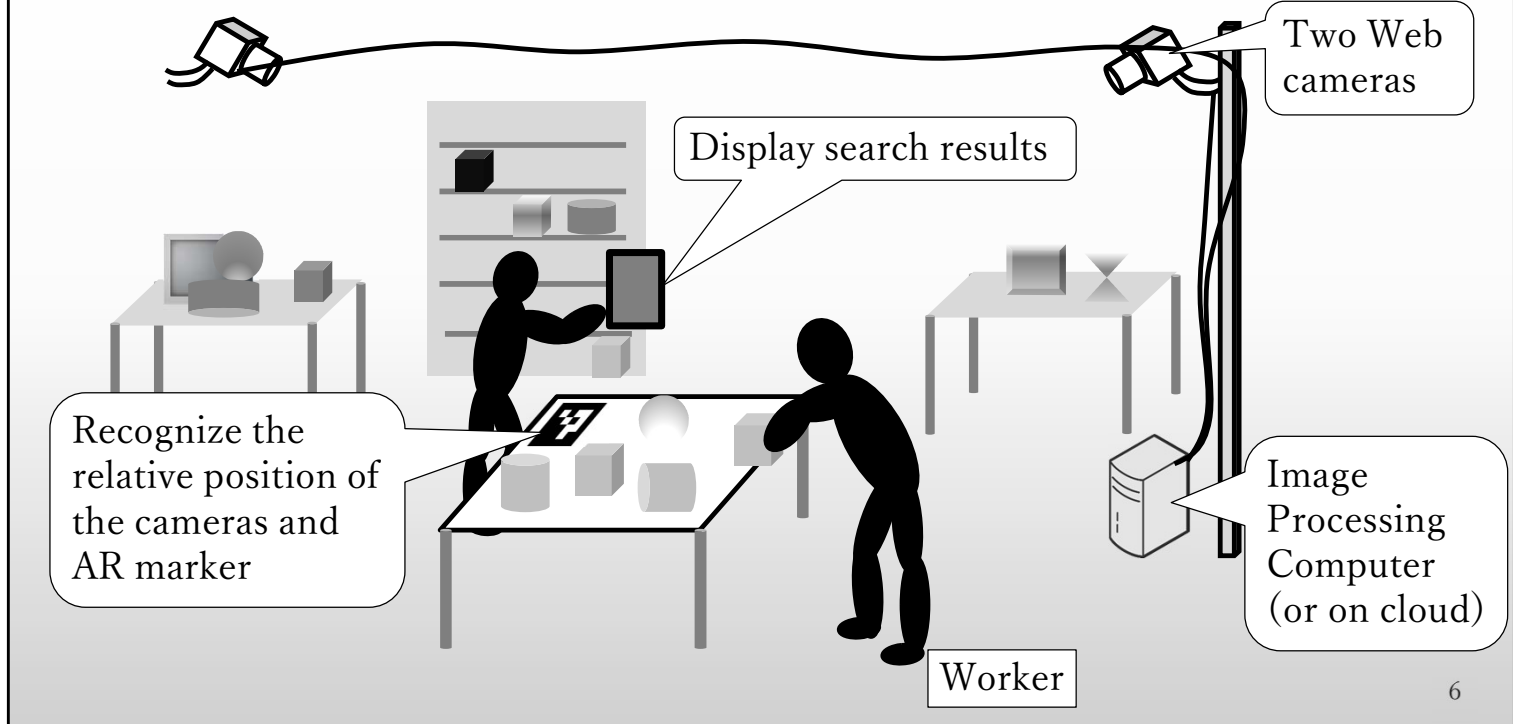
(a) PSS automatically collects and records data of various objects replaced in physical space without user involvement.

(b) PSS labels the recorded data of displaced objects with search tags such as "feature clustering numbers" based on the feature values, the colors they contain, and the date and time of movement.

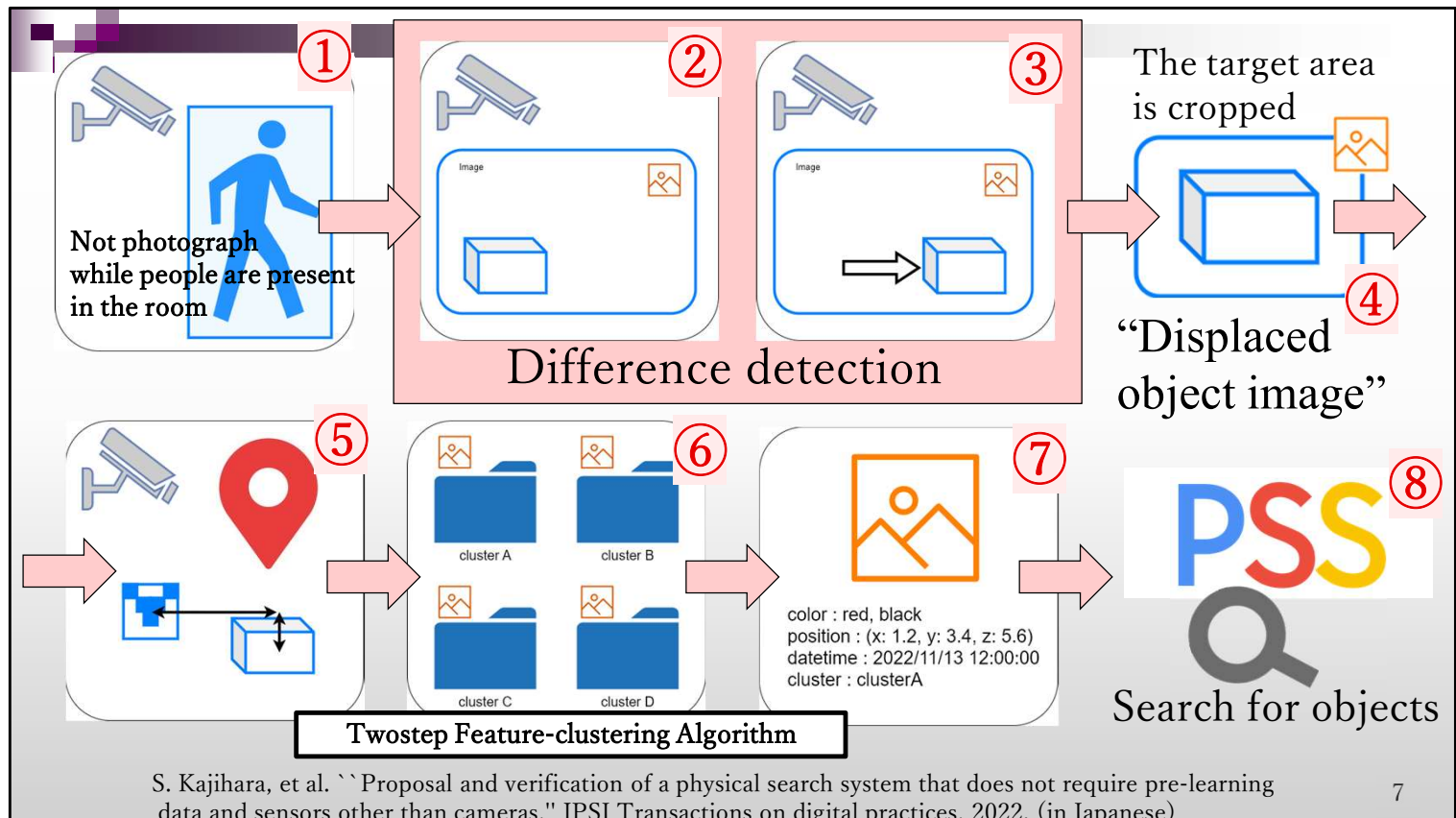
(c) Users can use PSS to search for a specific object (e.g., their own scissors) or similar objects (e.g., others' scissors). Search results displayed in PSS are information at a certain point in the past, and information at the present time is confirmed at the displayed location.

These characteristics are similar to those of google search in cyberspace.

The construction of PSS hardware



PSS is implemented in the environment shown in the figure. Two or more cameras are installed and fixed in a room, and they can also be linked to each other in each room. The cameras are asked to recognize AR markers of known sizes and to recognize the relative positions of the cameras and the AR markers. In this figure, the image processing computer is located in the room, but it can also be located in the cloud.



This slide shows the procedure of PSS.

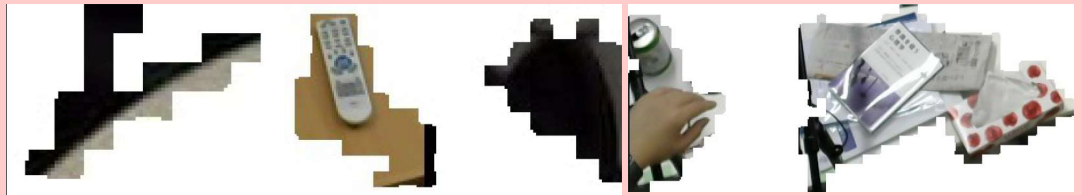
1. PSS does not photograph while people are present in the room.
2. When the people disappear in the room, the cameras photograph images in the room.
3. The image at a certain time is then compared at the pixel level with the image photographed at a previous time.
4. When a pixel with a difference of a certain standard or more is detected, it is determined that something has been displaced. The target area is cropped. The cropped image is called "displaced object image."
5. If the displaced object is captured by more than one camera, the 3D position of the displaced object is calculated.
6. The collected displaced object images are periodically clustered by twostep feature-clustering algorithm (TFA) to make them searchable by similar objects.
7. Each image stored in each cluster has information for retrieval: color, the date, time, location the image was taken.
8. The user can combine object clustering information to search for objects as needed.

Displaced object images


Success




Error







Some noisy images remain in a few clusters

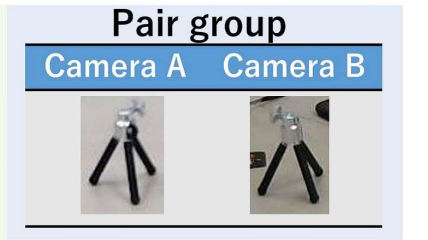
8

This figure shows examples of successful and unsuccessful cropped displaced object images. Noise images were generated due to light reflections and other effects.

Linking Method (LM)



Camera A	Camera B	Distance
		11
		21
		25
		28



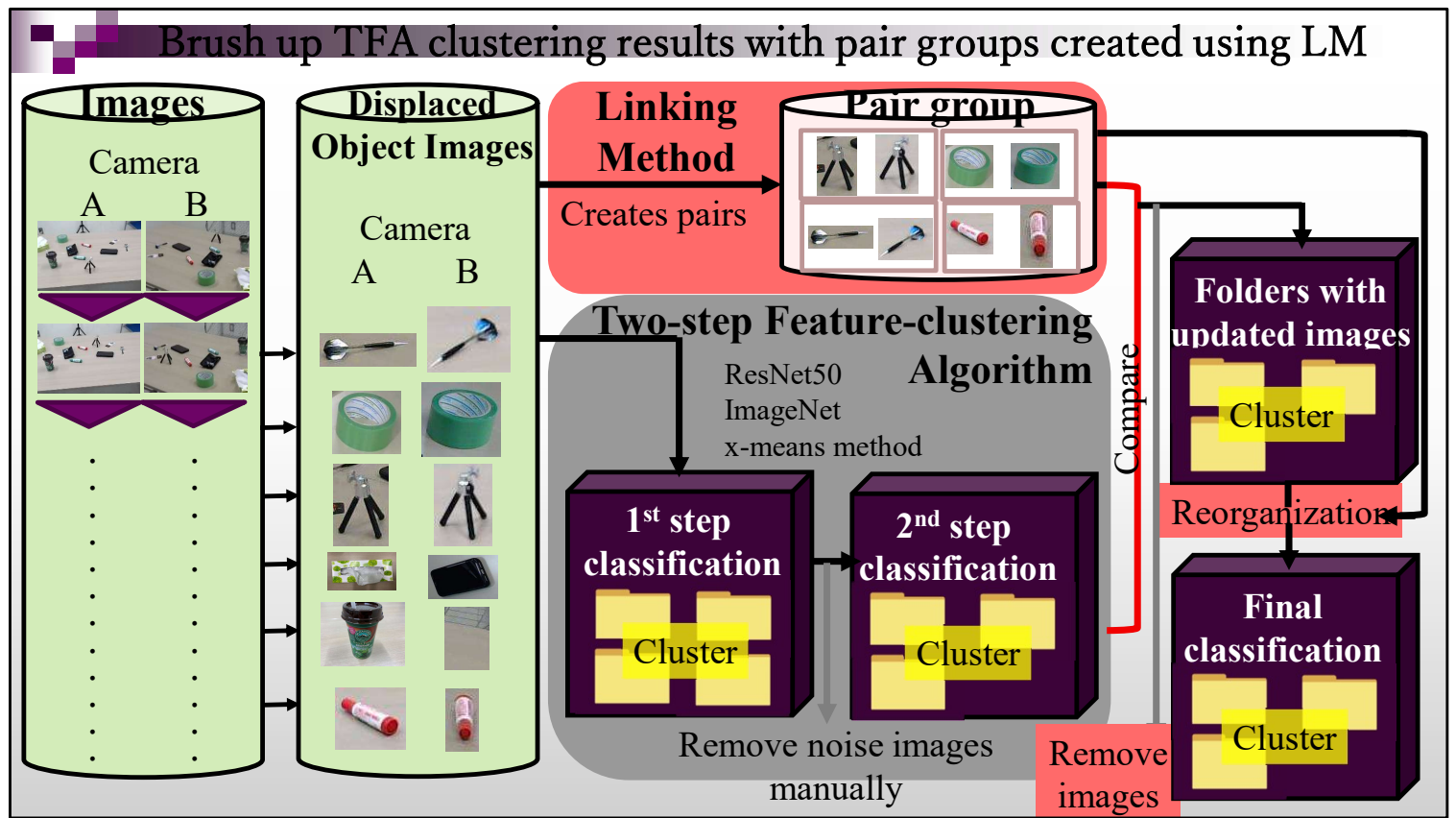
The left image has been paired with the other image.

The distance values exceed 23.

A2 and B2 are considered as noisy images

Therefore, in this paper, we propose and apply a linking method (LM) to improve the accuracy of the TFA clustering. For example, there are four displaced object images, A1 to B2. The A1 and A2 were photographed by Camera A as well as the B1 and B2 were photographed by Camera B. First, the similarity (distance) between A1 and B1, B2 and between A2 and B1, B2 are calculated. The smaller the distance between images, the higher the degree of image similarity. Namely, the distance between identical images is 0.

The A1 and B1, which have the smallest distance (11) among the remaining pairs, are considered as a pair. The image that is paired with another image is excluded from the candidate images for the other pairs. In addition, combinations with distance values exceeding 23 are not considered pairs. A gathering of the pairs is called a "pair group." In contrast, the A2 and B2 are considered as noisy images. Then, the images are removed.



This figure shows that the displaced object images are updated by comparing the TFA results with that of LM. The displaced object images in the clusters (TFA) that do not overlap with the displaced object images of the pair group (LM) are deleted.

Experiment

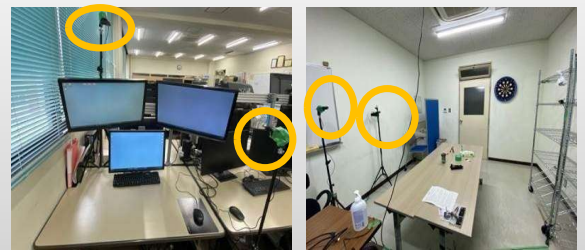
Compare the accuracy of clustering between **the combination of LM with TFA** and **TFA alone**.

1. Move one of the objects on the table
 2. Move it beyond the ranges of the cameras.
- Move each object 10 times in two conditions, Room C and D.
Cluster displaced object images in two Conditions, LM with TFA and TFA.



Ten kinds of objects

Two cameras were located



Room C

Room D

11

The objects were a red pen, a green pen, a smartphone tripod, a box of tissues, a cup of coffee, a black smartphone, a box of darts, a dart, a plastic bag of replacement dart feathers, and gum tape. Two cameras were located so that the entire table could be photographed from two different directions.

During the experiment, one of the authors moved one of the objects on the table and then moved beyond the ranges of the cameras. This method was applied to the ten objects. He moved each object 10 times in two conditions, Room C and D.

The PSS created clusters in two conditions, the combination of LM with TFA and TFA alone. The accuracy of the clustering is indicated in precision values, recall values, and F-measures. All displaced object images showing one of the ten objects are regarded as an "actual positive." The cluster in which the most images is included is considered to be a correct cluster, and the displaced object images of the correct cluster are regarded as a "predicted positive." In the predicted positive images, the actual positive images are considered to be true positives (TP), and the others are considered false positives (FP). In the actual positive images, the images that are not in the correct cluster are considered false negatives (FN).

Displaced objects	Recall		Precision		F-measure	
	TFA	LM with TFA	TA	LM with TFA	TFA	LM with TFA
Red pen	0.50	0.53	0.63	0.64	0.56	0.58
Green pen	0.38	0.43	0.47	0.43	0.42	0.43
Tripod	0.93	1.00	1.00	0.95	0.96	0.98
Box of tissues	0.49	0.63	1.00	1.00	0.65	0.77
Cup of coffee	0.91	1.00	1.00	0.94	0.95	0.97
Smartphone	0.42	0.58	0.23	0.29	0.30	0.39
Box of darts	0.44	0.69	0.30	0.30	0.36	0.42
Dart	0.75	0.75	0.24	0.27	0.36	0.40
Plastic bag	0.66	0.63	0.48	0.57	0.56	0.60
Gum tape	0.76	0.93	1.00	1.00	0.86	0.96
Average	0.62	0.72	0.64	0.64	0.60	0.65

12

This table shows the recall values, precision values, and F-measures to compare the results of TFA and LM with TFA conditions. For eight out of ten objects, the recall values were higher in the LM with TFA condition than in the TFA alone condition. In particular, the recall value of the gum tape under the LM with TFA condition was improved by 17%, compared to the TFA alone condition, and it became even closer to 100%. For all objects, the F-measures were higher in the LM with TFA condition than in the TFA alone condition. However, the precision values in the LM with TFA condition were almost the same as in the TFA alone condition. The displaced object images, the smartphone, the box of darts, and the dart remained around 30%.

Discussion

- Gum tape was groped into
 - Four clusters (TFA)
 - Some images missed a part of the gum tape
 - One cluster (LM with TFA)
- Difficult to group red and green pen into one cluster
 - An algorithm using color features should be applied
- Smartphone tripod and cup were paired in the LM
 - Something similar



13

The results showed that LM with TFA improved the clustering over TFA alone. As an example of improvement, the images of the gum tape were grouped into four clusters in the TFA alone condition because there were some images that missed a part of the gum tape. On the other hand, in the LM with TFA condition, most images of gum tape could be grouped into one cluster.

The F-measures for the red and green pens were not high, even in the LM with TFA condition. Since the shapes of these pens are similar, it was difficult to group them into one cluster using these algorithms. An algorithm using color features should be applied to such objects.

For the smartphone tripod and the cup of coffee, the precision values in the LM with TFA condition were lower than in the TFA alone condition. In the LM process, contrary to our expectations, the images of the smartphone tripod and the cup of coffee were paired with noisy images. Something similar to these objects was photographed as noisy images.

Conclusion

- Proposed a linking method (LM) that the images photographed simultaneously by two cameras were paired according to the images' similarity.
- Among the images in the cluster created by TFA alone, only the images paired using LM were left.
- The clustering accuracy was improved.
- The system is currently under construction as a system connected to a general network camera and running on a container in the cloud.

14

We proposed a linking method (LM). The images photographed simultaneously by two cameras before displaced objects were cropped out were paired according to the images' similarity. This method cannot allow most noisy images to be paired with other images. Finally, among the images in the cluster created by TFA alone, only the images paired using LM were left. As a result, the clustering accuracy was improved.

The system is currently under construction as a system connected to a general network camera and running on a container in the cloud.