# Online Object Recognition Using CNN-based Algorithm on High-speed Camera Imaging

## Framework for fast and robust high-speed camera object recognition based on population data cleansing and data ensemble

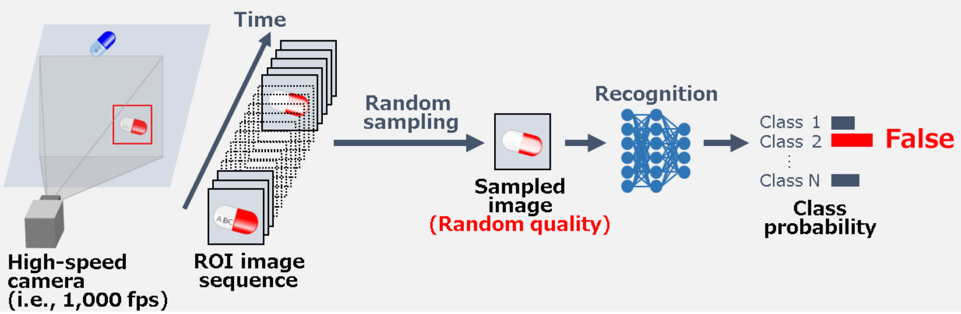
Shigeaki Namiki[1*], Keiko Yokoyama[1*], Shoji Yachida[1], Takashi Shibata[1†], Hiroyoshi Miyano[1], Masatoshi Ishikawa[2]

1. NEC, 2. Information Technology Center, The University of Tokyo *…Equally contributed †…this author now belongs to NTT

## Introduction

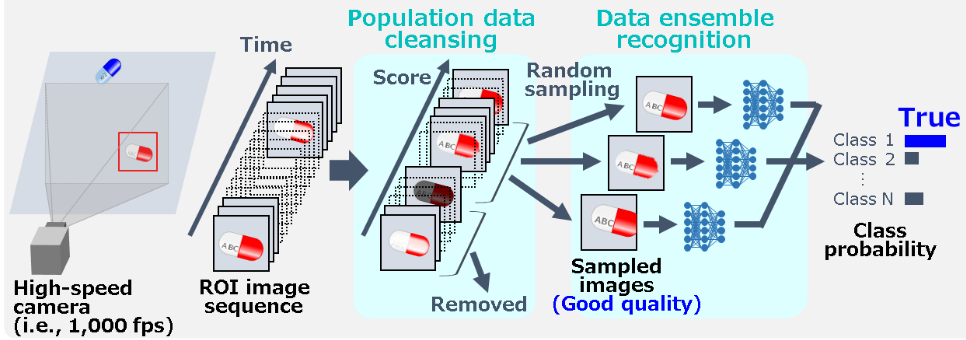
**Background:** High-speed camera has high temporal information density and low latency, which make fast moving object tracking and controlling easier. How about recognizing?
**Applications:** mass production lines, autonomous vehicles, etc.
**Problem:** low latency vs. high accuracy with temporally dense images.

Naive approach: random sampling. The accuracy depends on the quality of ROI images.




## Proposed Framework

1. **Population data cleansing** based on the recognizability score
   -> Remove low quality ROI images so as not to sample them.
2. **Data ensemble recognition** with a single light-weight CNN model
   -> more accurate, and more stable.




## Details of proposed framework

### 1. Population data cleansing:

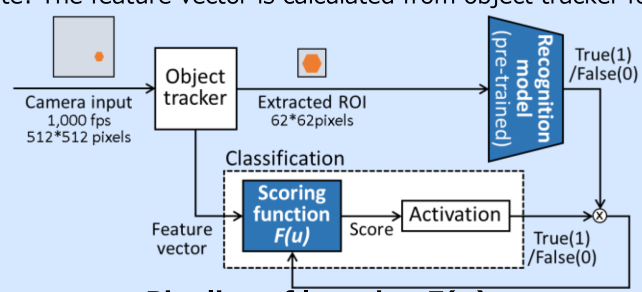
**Purpose:** Removes false ROI images, which recognition model yields as wrong labels.
**Limitations:** Need to be simple and fast to keep up with frame rate,
**Method:**

1. Label ROI images as true/false by pre-trained CNN used in Data Ensemble.
2. Train a simple linear classification model(Scoring functionF(u)) to predict the label using such as SVM or LDA.

$$F(\boldsymbol{u}) := \boldsymbol{w}^T \cdot \boldsymbol{u} + w_0 = 0$$

$w^T$: weight vector,
$u$: feature vector,
$w_0$: offset vector.

Note: The feature vector is calculated from object tracker for low latency.




**Pipeline of learning F(u)**

3. When testing, predict the scores and remove low score ROI images as low quality.

### 2. Data Ensemble recognition:

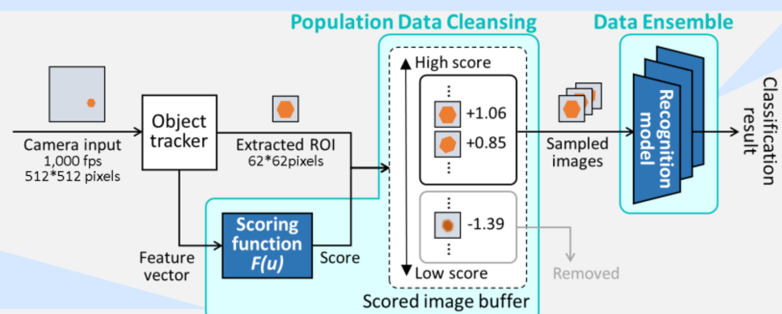
**Purpose:** Improve and stabilize the recognition accuracy
**Limitations:** Great model with a high accuracy cannot be used because of high latency
**Method:**

1. Construct light-weight model by decreasing the layers of existing CNN model.
2. Input multiple sampled images into the model and aggregate its outputs(C).

$$H^j(x_i, \ldots, x_{i+N}) = \frac{1}{N} \sum_{k=i}^{i+N} h^j(x_k),$$

$$C_i = \arg\max_j H^j(x_i, \ldots, x_{i+N})$$

$h^j$: $j-th$ class probability,
$x_i$: $i-th$ ROI image,
$H^j$: $j-th$ aggregated class probability,
$C_i$: predicted class for the sequence.




**Pipeline of our framework**

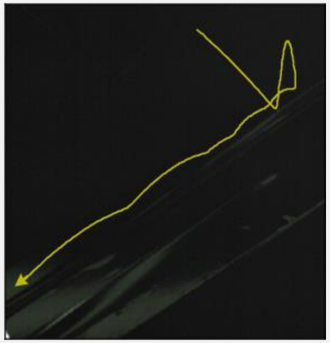
# Online Object Recognition Using CNN-based Algorithm on High-speed Camera Imaging

## Framework for fast and robust high-speed camera object recognition based on population data cleansing and data ensemble

Shigeaki Namiki[1*], Keiko Yokoyama[1*], Shoji Yachida[1], Takashi Shibata[1†], Hiroyoshi Miyano[1], Masatoshi Ishikawa[2]
1. NEC, 2. Information Technology Center, The University of Tokyo *…Equally contributed †…this author now belongs to NTT

## Dataset

- We construct A novel dataset inspired by the visual inspection in the mass production.
  - ✓ recorded by high-frame-rate camera (1000fps)
  - ✓ Target objects moving at high speed
  - ✓ Annotated with object categories

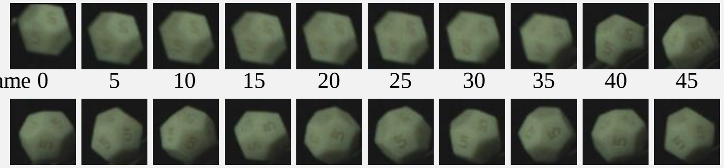


Background image.
the object bounces and slides on a slope.

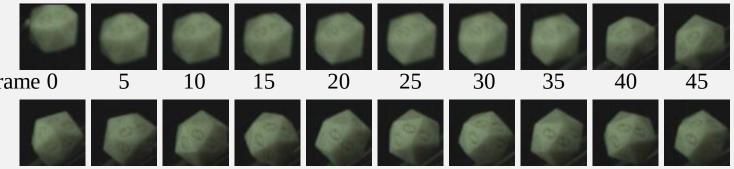
| | |
|---|---|
| Purpose | Object recognition |
| Label type | Object category |
| Target | 1-cm-diameter blocks |
| Resolution | 62 × 62 (Extracted from 512 × 512 images) |
| Frames per second (FPS) | 1,000 |
| Num. of classes | 20 (2 shapes, 10 characters) |

Specifications of the dataset

12-hedron, Character “5”




20-hedron, Character “0”




Examples of sequential ROI images in the dataset (every five frames)

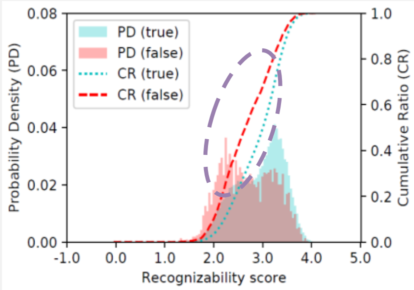
## Conclusions

- Proposed a novel object recognition framework for real-time applications with high-speed camera imaging
- Constructed A novel high-frame-rate video dataset for visual inspection
- Enabled CNN-based recognition against high-frame-rate time-series data in real time
- Showed More effective than existing approaches

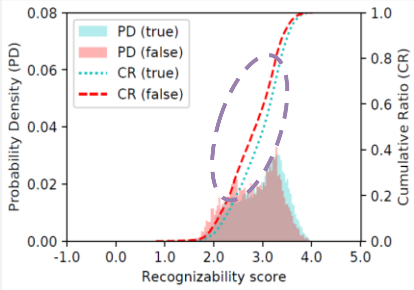
**This work is based on results obtained from a project commissioned by the New Energy and Industrial Technology Development Organization (NEDO).**

## Experiment

### 1. Improving recognizability by Population Data Cleansing(PDC):



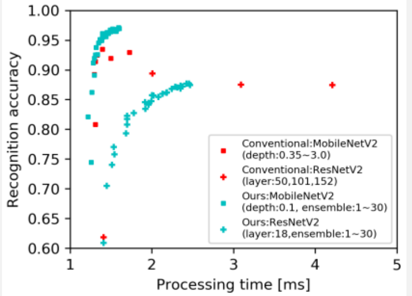

(a) training data



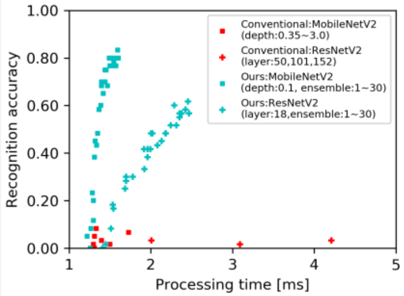

(b) test data

The lower recognizability score is, the higher the chances to remove false ROI images is.

### 2. Improving and Stabilizing Recognition Accuracy by Data Ensemble



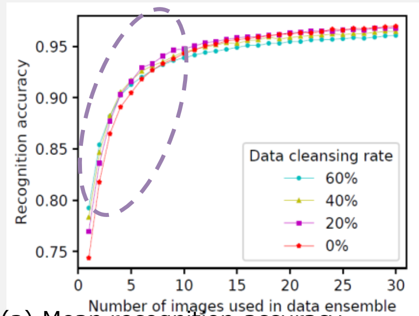

(a) Mean accuracy vs. processing time



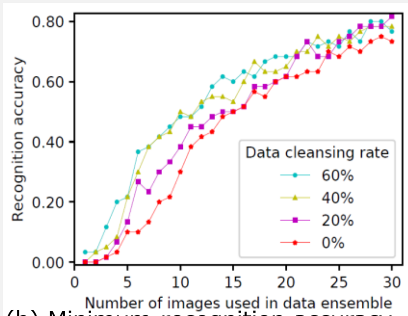

(b) Minimum accuracy vs. processing time

Ours is generally more accurate and definitely more stable than the conventional method.

### 3. Combination of the PDC and Data ensemble




(a) Mean recognition accuracy




(b) Minimum recognition accuracy

PDC keeps or slightly improves and stably suppresses the repeatability errors.