1. System Description

Our computing environment

Limited CPU and memory resources

Two off-the-shelf computers + GPUs (Titan Black)

Instead of using local features or motion features, 6 different CNNs were used.

[Step 1] Feature extraction with CNNs

(1) ImageNet
- Trained with the ImageNet dataset (1.2 million images and 1,000 categories)
- Provided with the Caffe (CNN) library

(2) Finetune
- Created with finetuning ImageNet model for TRECVID SIN task
- 1 million keyframe images
- 346 concepts (9 of units in the output layer: 346)

[Step 2] Feature pooling

Multiple frames from a shot

We selected a maximum of 10 frames from a shot at regular intervals.

[Step 3] Classification with SVMs

[Step 4] Classifier fusion

- Waseda4: Fusion weight of 2 for ImageNet, Finetune, Places and Hybrid models
- Fusion weight of 1 for Hybrid and Gradient models

- Waseda3: Fusion weight were optimized to improve the mAP of 30 concepts.
- Waseda2: Fusion weight were optimized to improve the mAP of 80 concepts.
- Waseda1: Fusion weight were optimized to improve the average precision of each concept.

2. Results of Submitted Runs

- Our 2015 submissions ranked between 5 and 8 in a total of 86 runs.
- Our best run ranked 2nd among all participants.

Comparison of Waseda runs with the runs of other teams on TRECVID 2015.

Scores from the following 3 scores were combined.
- Original images used for both training and testing.
- Both original and flipped images used for training, but only original images used for testing.
- Both original and flipped images used for training, and only flipped images used for testing.

Waseda1: 30.86
Waseda2: 30.73
Waseda3: 30.69
Waseda4: 30.69

Waseda at TRECVID 2015 Semantic Indexing
Kazuya UEKI, Tetsunori KOBAYASHI (Waseda University)