Extracting Spatio-Temporal Local Features Considering Consecutiveness of Motions

Akitsugu Noguchi and Keiji Yanai
The University of Electro-Communications, Tokyo, Japan

Background

◆Number of videos is increasing rapidly

- Web video like Youtube

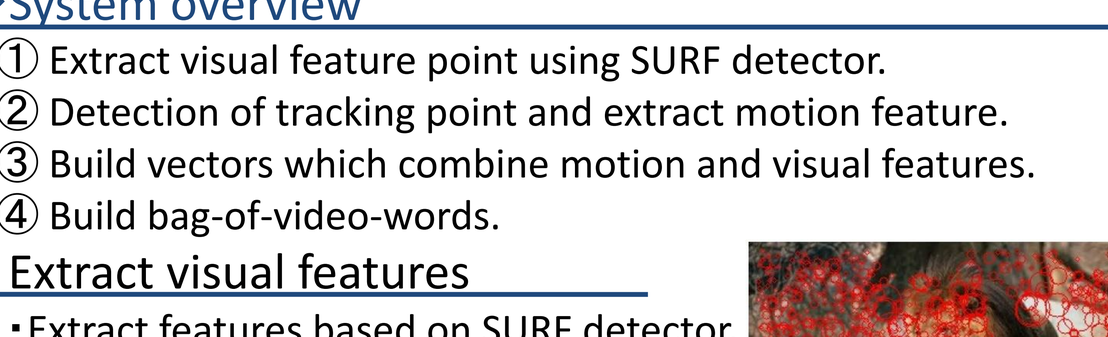
◆Video search is very difficult problem

- Text based search is utilized to search web for videos.
- It is difficult to search video which user want to watch only using text based search.

◆Content based video analysis will be needed

- Recently spatio-temporal feature has been proposed
- To investigate vast amount of video data, speed up technique is very important.

Past work



◆Existing method

- 3D Harris corner detector(Laptev et al. 2008)
- 2D-Gaussian filter and 1D-Gabor filter(Dollar et al. 2005)

◆Issue

- Extracting features from a whole cuboid costs much in term of computation.
- Is it necessary to extract features from whole cuboid?

◆In our work

- We want to extract features more fast and efficiently.
- We describe feature with a point and local track.
- Taking advance its high-speed performance, we apply to large scale web video search.

Proposed method

◆System overview

① Extract visual feature point using SURF detector.
② Detection of tracking point and extract motion feature.
③ Build vectors which combine motion and visual features.
④ Build bag-of-video-words.

1. Extract visual features

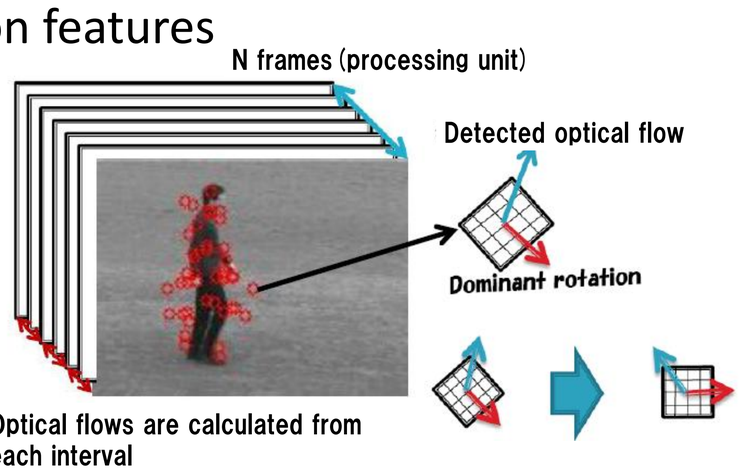
- Extract features based on SURF detector. (figure shows extracted points)

2. Detection of tracking points

- Points without motion are not suitable for spatio-temporal feature.
- In-motion points are selected by optical flow analysis.

Extraction of motion features

- Extract following N-frames as the processing unit.
- Divide N-frames into some interval, and optical flows are calculated from each interval.



- To make features more robust about rotation, rotate optical flow along the dominant direction of visual feature.

3.4. Build vector

- Concatenate visual and motion features into one vector with weight w.
- Build bag-of-video-words

Experiments

◆Human action classification

- Classify human action by SVM
- Dataset
  - KTH dataset which contains 6 motions
  - Each motion contains 100 videos
  - Multi-class classification with 5-fold-cross-validation
- Evaluate the following four combination of the feature
  - visual+motion+rotation(VMR)
  - visual+motion(VM)
  - visual only(V)
  - motion only(M)

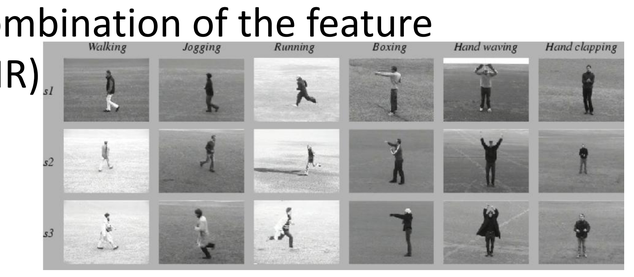


Table 1. Confusion matrix for VMR

	walking	running	jogging	boxing	waving	clapping
walking	0.94	0.02	0.03	0.01	0	0
running	0.02	0.76	0.22	0	0	0
jogging	0.04	0.15	0.81	0	0	0
boxing	0.01	0	0	0.91	0.02	0.07
waving	0	0	0	0.04	0.9	0.06
clapping	0	0	0	0.1	0.03	0.86

Table 2. Confusion matrix for V

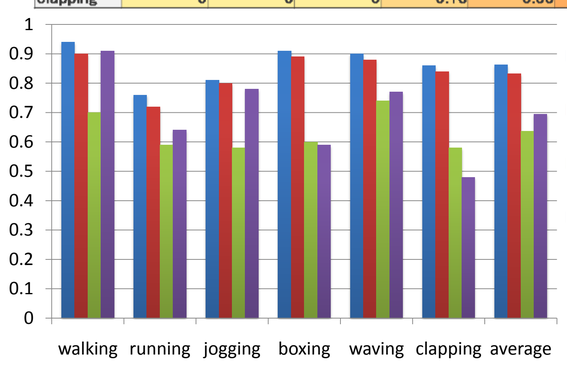
	walking	running	jogging	boxing	waving	clapping
walking	0.7	0.13	0.16	0.01	0	0
running	0.1	0.59	0.21	0	0	0
jogging	0.12	0.29	0.58	0	0	0.01
boxing	0.13	0.13	0.1	0.6	0.03	0.01
waving	0.03	0.09	0.01	0.05	0.74	0.08
clapping	0.04	0.05	0.02	0.06	0.25	0.58

Table 3. Confusion matrix for M

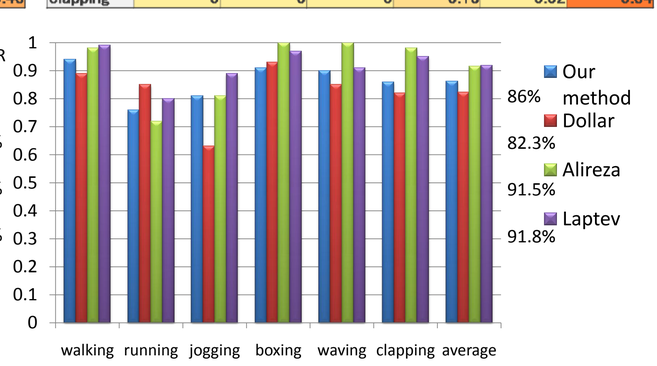
	walking	running	jogging	boxing	waving	clapping
walking	0.91	0	0.06	0.03	0	0
running	0	0.64	0.3	0	0.02	0.04
jogging	0.04	0.13	0.78	0.02	0.03	0
boxing	0.01	0	0	0.59	0.32	0.08
waving	0	0	0.01	0.17	0.77	0.05
clapping	0	0	0	0.18	0.33	0.48

Table 4. Confusion matrix for VM

	walking	running	jogging	boxing	waving	clapping
walking	0.9	0.01	0.07	0.01	0	0
running	0.01	0.72	0.27	0	0	0
jogging	0.01	0.18	0.8	0.01	0	0
boxing	0	0	0	0.89	0	0.11
waving	0	0	0	0.06	0.88	0.06
clapping	0	0	0	0.13	0.02	0.84



Result by four types of combination of feature



Comparison with other result by the state-of-art

◆Web video clustering

- Collect 100 soccer videos from Youtube
- Divide Each video into shots
- Extract features from each shot
- Classify web video shot by k-means clustering

Result of web video shot clustering per single video

Result of all web video shot clustering

Future work

◆To improve this feature

Detection of camera motion

◆Vast amount of web video shot clustering

Collect more than 1000 video
Clustering not only soccer video but many kinds of videos