Slow Feature Subspace for Action Recognition

05-14

Suzana Rita Alves Beleza[1], Kazuhiro Fukui[1,2]

[1] *Graduate School of Systems and Information Engineering, University of Tsukuba, Japan*

[2] *Center for Artificial Intelligence Research, Tsukuba, Japan*

(1) Motivation and objective

- Human action recognition has received renewed attention due to new applications in surveillance, human-machine interaction and assistive technologies
- The goal of my research is to recognize actions using a combination of **subspace-based methods** and **slow feature analysis (SFA)**[1].
- **PCA-SFA**[2], a variant of SFA, produces a video descriptor from the projection vectors of the SF space, that also contain the temporal information
- Therefore, in my research, we generate a video descriptor from an input video using the projection vectors of PCA-SFA, and then transform this descriptor into a subspace by applying PCA.
- This new **slow feature subspace**[3] includes temporal dynamic information.
- And thus, we can classify actions by simply calculating the similarity between their subspaces.

(2) Flow of the proposed framework

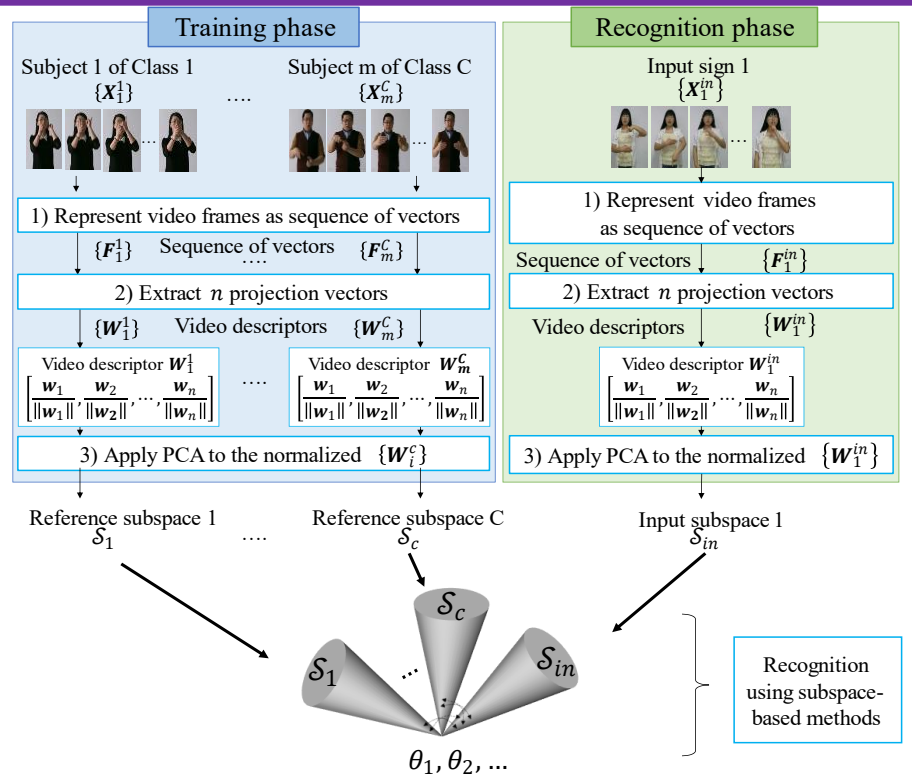

Fig 1. The process flow of the proposed framework.

(3) Databases

- KTH action[4] dataset has videos with six classes. The actions were performed by 25 individuals on different scenarios: outdoors, outdoors with zoom variation, outdoors with different clothes and indoors. Each has four samples.

Fig 2. Examples of KTH action dataset

- Chinese sign language dataset (isolated SLR500)[5]: it consists of 500 signs, each performed by 50 subjects. We randomly sampled 10 signs, considering only the first 50 frames of each video. We also extracted their CNN features using VGG-19.

Fig 3. Examples of isolated SLR500

(4) Experimental results

Data	MSM	CMSM	KMSM	KCMSM
KTH raw	67.3±2.6	67.1±3.9	67.6±2.2	65.8±2.1
Proj Vec(5)	67.0±1.7	65.7±2.1	71.4±1.4	72.0±1.9
Proj Vec(8)	68.8±1.7	67.6±2.1	72.1±1.1	72.2±2.0
Proj Vec(10)	**71.2±2.3**	**69.7±2.1**	**73.1±1.2**	**73.7±2.3**

Table 1. Classification accuracy (%) for KTH action.

Data	MSM	CMSM	KMSM	KCMSM	Softmax
Isolated SLR500	-	-	-	-	30.7±3.9
CNN features	59.8±5.5	60.4±4.4	67.6±5.9	59.3±4.0	-
Proj Vec(5)	71.4±3.9	64.7±7.1	79.5±3.4	**78.6±4.0**	-
Proj Vec(8)	70.2±3.9	66.8±4.5	**80.7±4.4**	77.1±2.2	-
Proj Vec(10)	**73.0±4.1**	**67.5±4.2**	80.6±3.1	77.9±4.0	-
Proj Vec(20)	67.9±4.2	66.6±5.0	72.2±2.6	67.6±2.6	-
Proj Vec(50)	53.9±4.5	52.3±6.7	68.7±3.3	67.1±4.2	-

Table 2. Classification accuracy (%) of fine-tuned VGG-19 features for Isolated SLR500.

(5) Conclusions

- We proposed a framework for action recognition jointly using subspace-based methods and SFA
- Our framework could improve the action recognition independently of using raw images or CNN features
- Our framework also outperformed the accuracy results of the fine-tuned VGG-19 model

(6) References

[1] Wiskott, L., Sejnowski, T.J.: Slow feature analysis: unsupervised learning of invariances. Neural Computation 14(4), 715-770. 2002.

[2] Takumi Kobayashi. Feature sequence representation via slow feature analysis for action classification. In Proceedings of the British Machine Vision Conference (BMVC), pages125.1–125.13, 2017.

[3] Beleza, Suzana R. A, and Fukui, Kazuhiro. "Slow Feature Subspace for Action Recognition". *International Conference on Pattern Recognition (ICPR)* 2020 Workshops. Springer, in press.

[4] https://www.csc.kth.se/cvap/actions/

[5] http://home.ustc.edu.cn/~pjh/openresources/cslr-dataset-2015/index.html