

Team name	CUHK-SWJTU
Team leader name	Limin Wang
Team leader address, phone number and email	Room 703, Ho Sin-Hang Engineering Building The Chinese University of Hong Kong Shatin, N.T. , Hong Kong E-mail: <a href="mailto:07wanglimin@gmail.com">07wanglimin@gmail.com</a> ; Phone: (+86) 13144825084
Rest of team members	Xiaojiang Peng
Team website URL (if any)	Personal homepage: Limin Wang: <a href="http://wanglimin.github.io">http://wanglimin.github.io</a> ; Xiaojiang Peng: <a href="http://pengxj.github.io">http://pengxj.github.io</a> .

Title of the contribution	Gesture spotting by using color features based on spatiotemporal segmentation
General method description	<ol style="list-style-type: none"><li data-bbox="600 232 1875 332">1. We firstly segment each long video stream into short clips by analyzing the trajectories of hand joints.</li><li data-bbox="600 347 1856 504">2. For each video clip, we extract the color features, namely dense trajectories with descriptors of HOG, HOF, MBHx, and MBHy.</li><li data-bbox="600 518 1731 675">3. For dense trajectories, we use Fisher vector encoding method and VLAD encoding method with power+L2-normalization. We then train a linear SVM classifier.</li><li data-bbox="600 689 1779 789">4. We use post-processing techniques to delete the clip of background class.</li></ol>
References	<ol style="list-style-type: none"><li data-bbox="600 903 1856 989">1. H. Wang, A Klaser, C Schmid and CL Liu, Action recognition by dense trajectories, in CVPR 2010.</li><li data-bbox="600 1003 1798 1146">2. J. Sanchez, F. Perronnin, T. Mensink, and J. Verbeek, Image Classification with the Fisher Vector: Theory and Practice, in IJCV 2013.</li><li data-bbox="600 1160 1789 1246">3. Jegou H., Perronnin F., Douze, M., Sanchez, J., Aggregating Local Image Descriptors into Compact Codes, in PAMI 2012.</li></ol>

Describe data preprocessing techniques applied (if any)	No.
Describe features used or data representation model (if any)	Dense trajectories.
Data modalities used, i.e. depth, rgb, skeleton... (if any)	rgb.
Fusion strategy applied (if any)	No.
Dimensionality reduction technique applied (if any)	PCA dimension reduction.

Temporal clustering approach (if any)	No.
Temporal segmentation approach (if any)	Segment the video into clips using skeleton and segmentation mask.
Gesture representation approach (if any)	Fisher vector and VLAD of dense trajectories.
Classifier used (if any)	SVM
Large scale strategy (if any)	No.

Transfer learning strategy (if any)	No.
Temporal coherence and/or tracking approach considered (if any)	No.
Other technique/strategy used not included in previous items (if any)	No.
Method complexity analysis	Training: about 20 hours on PC with CPU (Intel i7-4770), RAM (32G)  Testing: about 10 hours on PC with CPU (Intel i7-4770), RAM (32G)

**Qualitative advantages of the proposed solution**

**Only one kind of modality is used. The method is simple and efficient.**

Results of the comparison to other approaches (if any)

No.

Novelty degree of the solution and if it has been previously published

We propose a simple yet effective method for gesture temporal spotting and recognition. Specifically, we design a temporal segmentation algorithm based on analysis of human hand joints. Then we resort to dense trajectory features and hybrid super vector encoding methods (FV+VLAD) to obtain a global representation for each video clip. Finally we use SVM classifier for recognition.

Language and implementation details (including platform, memory, parallelization requirements)	Matlab 2012 <sup>a</sup> ,on Window 7 with 64 bit; Vsual Studio 2010, with Opencv 2.4.5; Parallelization: matlab with 4 cores.
Human effort required for implementation, training and validation?	No.
Training/testing expended time?	Training: 20 hours; Testing: 9 hours
General comments and impressions of the challenge	Very good. Thanks for your organization. The schedule should be fixed just as the submission deadline of a conference.