

Team name	DonkeyBurger
Team leader name	Yi Tian
Team leader address, phone number and email	Institute of Information Science Beijing Jiaotong University, Beijing 100044, China +86 13810308506 11112062@bjtu.edu.cn
Rest of team members	Ruoyu Liu
Team website URL (if any)	

Title of the contribution	<ol style="list-style-type: none"> 1. Two phase of action recognition in image sequence 2. Mean value and variance of reconstruction error are used to segment the video sequence
General method description	<p>At first phase, the MHI[1] (Motion History Image) is extracted, and we use this descriptor to recognize some special actions such as walking, running, jumping and crouching.</p> <p>At second phase, the STIP[2] (Space Time Interest Points) is extracted. K-means is used to cluster dictionary, under which each point is represented as sequence of sparse codes[3]. After that, each frame of videos is described as the histogram of interest points. We use this descriptor to train dictionaries separately for each kind of single and inter actions and calculate the sparse codes[4]. Finally, the loss of reconstruction is used to judge the label of each frame.</p>
References	<p>[1] Davis J W. Hierarchical motion history images for recognizing human motion[C]//Detection and Recognition of Events in Video, 2001. Proceedings. IEEE Workshop on. IEEE, 2001: 39-46.</p> <p>[2] Laptev I. On space-time interest points[J]. International Journal of Computer Vision, 2005, 64(2-3): 107-123.</p> <p>[3] Lee H, Battle A, Raina R, et al. Efficient sparse coding algorithms[J]. Advances in neural information processing systems, 2007, 19: 801.</p> <p>[4] Wang J, Yang J, Yu K, et al. Locality-constrained linear coding for image classification[C]//Computer Vision and Pattern</p>

Describe data preprocessing techniques applied (if any)	Convert the videos into proper format to extract features. Firstly, we use 'xvid' encoder to recode videos for extracting MHI on MATLAB platform. Then we use 'H.263' encoder to recode videos for extracting STIP using OpenCV library.
Describe features used or data representation model (if any)	Global feature: MHI (Motion History Image) Local feature: STIP (Space Time Interest Points)
Dimensionality reduction technique applied (if any)	
Temporal clustering approach (if any)	K-means

Temporal segmentation approach (if any)	We segment the video frame by frame. Firstly, we calculate the reconstruction error of each frame under dictionaries of different actions. And then, mean value and variance are used to identify whether the frame belongs to 'unrecognizable' . After that, minimum reconstruction error criterion is used to classify these left 'recognizable' frame.
Gesture representation approach (if any)	
Classifier used (if any)	Sparse code is calculated for each frame of videos, and the reconstruction error under different action dictionary is used to classify.
Large scale strategy (if any)	

Transfer learning strategy (if any)	
Temporal coherence and/or tracking approach considered (if any)	Each frame is described as a cube which is constructed with its neighbor frames.
Compositional model used, i.e. pictorial structure (if any)	
Other technique/strategy used not included in previous items (if any)	Sparse coding and Locality-constrained linear coding
Method complexity analysis	<p>The first phase of method is fast, because of no need of training model.</p> <p>The second phase of method is slowly, because it needs to learn dictionaries and cluster.</p>

Qualitative advantages of the proposed solution

**Use first phase to pre-recognize some single actions and separate the frames of single and inter actions, in order to decrease the complex of action recognition task.
The reconstruction error is more discrimination for segmentation than some common classifiers, such as SVM.
Not only use the value of reconstruction error, but also the mean value and variance of it to segment the video sequence.**

Results of the comparison to other approaches (if any)

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

A little creativity.
It hasn't been previously published.

Language and implementation details (including platform, memory, parallelization requirements)	Matlab 2013a on Windows 8 4GB memory Single thread
Human effort required for implementation, training and validation?	Need to adjust some parameters.
Training/testing expended time?	Training time: about half an hour Testing time: less than 1 minute
General comments and impressions of the challenge	Some description of the schedule of the challenge is hard-understanding. Need a better and unit platform to submit code and prediction results.