

## Accepted Manuscript

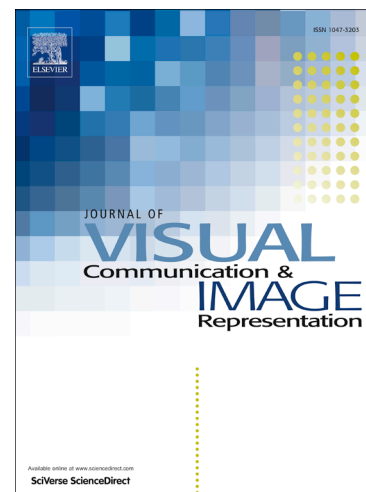
A novel recurrent hybrid network for feature fusion in action recognition

Sheng Yu, Yun Cheng, Li Xie, Zhiming Luo, Min Huang, Shaozi Li

PII: S1047-3203(17)30181-5  
DOI: <http://dx.doi.org/10.1016/j.jvcir.2017.09.007>  
Reference: YJVICI 2056

To appear in: *J. Vis. Commun. Image R.*

Received Date: 8 July 2017  
Accepted Date: 14 September 2017



Please cite this article as: S. Yu, Y. Cheng, L. Xie, Z. Luo, M. Huang, S. Li, A novel recurrent hybrid network for feature fusion in action recognition, *J. Vis. Commun. Image R.* (2017), doi: <http://dx.doi.org/10.1016/j.jvcir.2017.09.007>

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

# A novel recurrent hybrid network for feature fusion in action recognition

Sheng Yu<sup>a,b,c</sup>, Yun Cheng<sup>b</sup>, Li Xie<sup>b</sup>, Zhiming Luo<sup>a,c</sup>, Min Huang<sup>a,c</sup>,  
Shaozi Li<sup>a,c,\*</sup>

<sup>a</sup>*Department of Cognitive Science, Xiamen University, Xiamen, Fujian, 361005, China*

<sup>b</sup>*School of Information, Hunan University of Humanities, Science and Technology, Loudi, Hunan, China*

<sup>c</sup>*Fujian Key Laboratory of the Brain-like Intelligent Systems, Xiamen, Fujian, China*

---

## Abstract

Action recognition in video is one of the most important and challenging tasks in computer vision. How to efficiently combine the spatial-temporal information to represent video plays a crucial role for action recognition. In this paper, a recurrent hybrid network architecture is designed for action recognition by fusing multi-source features: a two-stream CNNs for learning semantic features, a two-stream single-layer LSTM for learning long-term temporal feature, and an Improved Dense Trajectories (IDT) stream for learning short-term temporal motion feature. In order to mitigate the overfitting issue on small-scale dataset, a video data augmentation method is used to increase the amount of training data, as well as a two-step training strategy is adopted to train our recurrent hybrid network. Experiment results on two challenging datasets UCF-101 and HMDB-51 demonstrate that the proposed method can reach the state-of-the-art performance.

*Keywords:* Deep learning, action recognition, LSTM, CNNs, IDT

---

\*Corresponding author

*Email address:* szlig@xmu.edu.cn (Shaozi Li)

Download English Version:

<https://daneshyari.com/en/article/4969234>

Download Persian Version:

<https://daneshyari.com/article/4969234>

[Daneshyari.com](https://daneshyari.com)