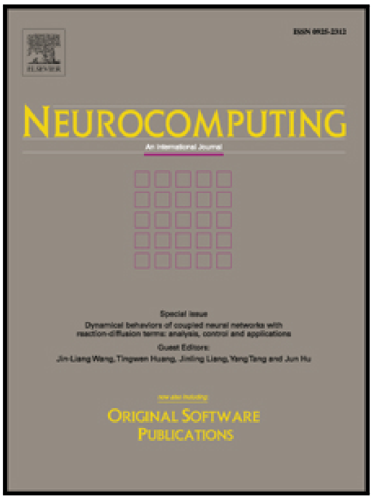
Communicated by Mingli Song



Multi-cue fusion for emotion recognition in the wild

Jingwei Yan, Wenming Zheng, Zhen Cui, Chuangao Tang,
Tong Zhang, Yuan Zong

PII:	S0925-2312(18)30486-7
DOI:	10.1016/j.neucom.2018.03.068
Reference:	NEUCOM 19519
To appear in:	*Neurocomputing*
Received date:	22 July 2017
Revised date:	1 March 2018
Accepted date:	28 March 2018

Multi-cue fusion for emotion recognition in the wild

Jingwei Yan[a], Wenming Zheng[a,*], Zhen Cui[b], Chuangao Tang[a], Tong Zhang[c], Yuan Zong[a]

[a]*Key Laboratory of Child Development and Learning Science of Ministry of Education, School of Biological Science and Medical Engineering, Southeast University, Nanjing 210096, China*
[b]*School of Computer Science and Engineering, Nanjing University of Science and Technology, Nanjing 210094, China*
[c]*School of Information Science and Engineering, Southeast University, Nanjing 210096, China*

Abstract

Emotion recognition has become a hot research topic in the past several years due to the large demand of this technology in many practical situations. One challenging task in this topic is to recognize emotion types in a given video clip collected in the wild. In order to solve this problem we propose a multi-cue fusion emotion recognition (MCFER) framework by modeling human emotions from three complementary cues, *i.e.*, facial texture, facial landmark action and audio signal, and then fusing them together. To capture the dynamic change of facial texture we employ a cascaded convolutional neutral network (CNN) and bidirectional recurrent neutral network (BRNN) architecture where facial image from each frame is first fed into CNN to extract high-level texture feature, and then the feature sequence is traversed into BRNN to learn the changes within it. Facial landmark action models the movement of facial muscles explicitly. SVM and CNN are deployed to explore the emotion related patterns in it. Audio signal is also modeled with CNN by extracting low-level acoustic features from segmented clips and then stacking them as an image-like matrix. We fuse these models at both feature level and decision level to further boost the overall performance. Experimental results on two challenging databases demonstrate

[*] Corresponding author
Email address: wenming_zheng@seu.edu.cn (Wenming Zheng)