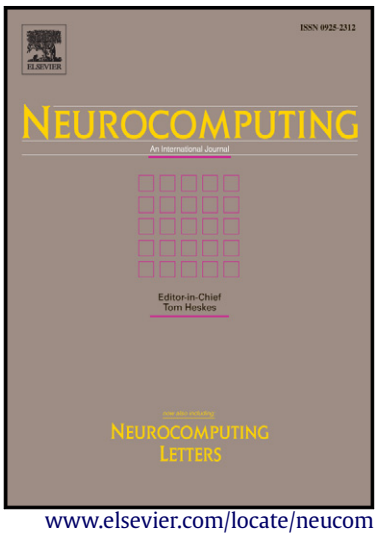
Author's Accepted Manuscript



www.elsevier.com/locate/neucom

Spatial and temporal Visual attention prediction in videos using eye movement data

Junwei Han, Liye Sun, Xintao Hu, Jungong Han, Ling Shao

PII: S0925-2312(14)00698-5
DOI: http://dx.doi.org/10.1016/j.neucom.2014.05.049
Reference: NEUCOM14255

To appear in: *Neurocomputing*

Received date: 3 August 2013
Revised date: 18 April 2014
Accepted date: 26 May 2014

Cite this article as: Junwei Han, Liye Sun, Xintao Hu, Jungong Han, Ling Shao, Spatial and temporal Visual attention prediction in videos using eye movement data, *Neurocomputing*, http://dx.doi.org/10.1016/j.neucom.2014.05.049

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 galley proof before it is published in its final citable 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.

ACCEPTED MANUSCRIPT

Spatial and temporal visual attention prediction in videos using eye movement data

Junwei Han[1], Liye Sun[1], Xintao Hu[1], Jungong Han[2*], Ling Shao[3]

[1]*School of Automation, Northwestern Polytechnical University, Xi'an, 710072, China*

[2]*Civolution Technology, Eindhoven, The Netherlands*

[3]*The University of Sheffield, Sheffield, S1 3JD, UK*

**Corresponding author Email: jungonghan77@gmail.com*

Abstract:

Visual attention detection in static images has achieved outstanding progress in recent years whereas much less effort has been devoted to learning visual attention in video sequences. In this paper, we propose a novel method to model spatial and temporal visual attention for videos respectively through learning from human gaze data. The spatial visual attention mainly predicts where viewers look in each video frame while the temporal visual attention measures which video frame is more likely to attract viewers' interest. Our underlying premise is that objects as well as their movements, instead of conventional contrast-related information, are major factors in dynamic scenes to drive visual attention. Firstly, the proposed models extract two types of bottom-up features derived from multi-scale object filter responses and spatiotemporal motion energy, respectively. Then, spatiotemporal gaze density and inter-observer gaze congruency are generated using a large collection of human-eye gaze data to form two training sets. Finally, prediction models of temporal visual attention and spatial visual attention are learned based on those two training sets and bottom-up features, respectively. Extensive evaluations on publicly available video benchmarks and applications in interestingness prediction of movie trailers demonstrate the effectiveness of the proposed work.

Key words: spatial visual attention, temporal visual attention, human gaze, spatiotemporal gaze density, fixation dispersion.

Accepted manuscript

Download English Version:

https://daneshyari.com/en/article/6866352

Download Persian Version:

https://daneshyari.com/article/6866352

Daneshyari.com