

Jianguo Zhang, Ling Shao, Lei Zhang, and Graeme A. Jones (Eds.)

Intelligent Video Event Analysis and Understanding

Studies in Computational Intelligence, Volume 332

Editor-in-Chief

Prof. Janusz Kacprzyk
Systems Research Institute
Polish Academy of Sciences
ul. Newelska 6
01-447 Warsaw
Poland
E-mail: kacprzyk@ibspan.waw.pl

Further volumes of this series can be found on our homepage: springer.com

Vol. 312. Patricia Melin, Janusz Kacprzyk, and Witold Pedrycz (Eds.)
Soft Computing for Recognition based on Biometrics, 2010
ISBN 978-3-642-15110-1

Vol. 313. Imre J. Rudas, János Fodor, and Janusz Kacprzyk (Eds.)
Computational Intelligence in Engineering, 2010
ISBN 978-3-642-15219-1

Vol. 314. Lorenzo Magnani, Walter Carnielli, and Claudio Pizzi (Eds.)
Model-Based Reasoning in Science and Technology, 2010
ISBN 978-3-642-15222-1

Vol. 315. Mohammad Essaaidi, Michele Malgeri, and Costin Badica (Eds.)
Intelligent Distributed Computing IV, 2010
ISBN 978-3-642-15210-8

Vol. 316. Philipp Wolfrum
Information Routing, Correspondence Finding, and Object Recognition in the Brain, 2010
ISBN 978-3-642-15253-5

Vol. 317. Roger Lee (Ed.)
Computer and Information Science 2010
ISBN 978-3-642-15404-1

Vol. 318. Oscar Castillo, Janusz Kacprzyk, and Witold Pedrycz (Eds.)
Soft Computing for Intelligent Control and Mobile Robotics, 2010
ISBN 978-3-642-15533-8

Vol. 319. Takayuki Ito, Minjie Zhang, Valentin Robu, Shaheen Fatima, Tokuro Matsuo, and Hirofumi Yamaki (Eds.)
Innovations in Agent-Based Complex Automated Negotiations, 2010
ISBN 978-3-642-15611-3

Vol. 320. xxx

Vol. 321. Dimitri Plemenos and Georgios Miaoulis (Eds.)
Intelligent Computer Graphics 2010
ISBN 978-3-642-15689-2

Vol. 322. Bruno Baruque and Emilio Corchado (Eds.)
Fusion Methods for Unsupervised Learning Ensembles, 2010
ISBN 978-3-642-16204-6

Vol. 323. Yingxu Wang, Du Zhang, and Witold Kinsner (Eds.)
Advances in Cognitive Informatics, 2010
ISBN 978-3-642-16082-0

Vol. 324. Alessandro Soro, Vargiu Eloisa, Giuliano Armano, and Gavino Paddeu (Eds.)
Information Retrieval and Mining in Distributed Environments, 2010
ISBN 978-3-642-16088-2

Vol. 325. Quan Bai and Naoki Fukuta (Eds.)
Advances in Practical Multi-Agent Systems, 2010
ISBN 978-3-642-16097-4

Vol. 326. Sheryl Brahmam and Lakhmi C. Jain (Eds.)
Advanced Computational Intelligence Paradigms in Healthcare 5, 2010
ISBN 978-3-642-16094-3

Vol. 327. Sławomir Wiak and Ewa Napieralska-Juszczak (Eds.)
Computational Methods for the Innovative Design of Electrical Devices, 2010
ISBN 978-3-642-16224-4

Vol. 328. Raoul Huys and Viktor K. Jirsa (Eds.)
Nonlinear Dynamics in Human Behavior, 2010
ISBN 978-3-642-16261-9

Vol. 329. Santi Caballé, Fatos Xhafa, and Ajith Abraham (Eds.)
Intelligent Networking, Collaborative Systems and Applications, 2010
ISBN 978-3-642-16792-8

Vol. 330. Steffen Rendle
Context-Aware Ranking with Factorization Models, 2010
ISBN 978-3-642-16897-0

Vol. 331. Athena Vakali and Lakhmi C. Jain (Eds.)
New Directions in Web Data Management 1, 2011
ISBN 978-3-642-17550-3

Vol. 332. Jianguo Zhang, Ling Shao, Lei Zhang, and Graeme A. Jones (Eds.)
Intelligent Video Event Analysis and Understanding, 2011
ISBN 978-3-642-17553-4

Jianguo Zhang, Ling Shao, Lei Zhang,
and Graeme A. Jones (Eds.)

Intelligent Video Event Analysis and Understanding

Dr. Jianguo Zhang

School of Computing

University of Dundee

Dundee DD1 4HN

Scotland, UK

E-mail: jgzhang@computing.dundee.ac.uk

<http://www.computing.dundee.ac.uk/staff/jgzhang/>

Dr. Ling Shao

Department of Electronic & Electrical Engineering

The University of Sheffield, Sheffield, S1 3JD, UK

E-mail: ling.shao@sheffield.ac.uk

<http://lshao.staff.shef.ac.uk/>

Dr. Lei Zhang

Lead Researcher, Microsoft Research Asia,

49 Zhichun Road, Beijing 100190, P.R. China

E-mail: leizhang@microsoft.com

<http://research.microsoft.com/en-us/people/leizhang/>

Prof. Graeme A. Jones

Digital Imaging Research Centre,

Faculty of Computing, Information Systems and Mathematics,

Kingston University, Penrhyn Road, Kingston upon Thames, Surrey KT1 2EE, UK

E-mail: g.jones@kingston.ac.uk

<http://cism.kingston.ac.uk/people/details.php?AuthorID=3>

ISBN 978-3-642-17553-4

e-ISBN 978-3-642-17554-1

DOI 10.1007/978-3-642-17554-1

Studies in Computational Intelligence

ISSN 1860-949X

© 2011 Springer-Verlag Berlin Heidelberg

This work is subject to copyright. All rights are reserved, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilm or in any other way, and storage in data banks. Duplication of this publication or parts thereof is permitted only under the provisions of the German Copyright Law of September 9, 1965, in its current version, and permission for use must always be obtained from Springer. Violations are liable to prosecution under the German Copyright Law.

The use of general descriptive names, registered names, trademarks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.

Typeset & Cover Design: Scientific Publishing Services Pvt. Ltd., Chennai, India.

Printed on acid-free paper

9 8 7 6 5 4 3 2 1

springer.com

Preface

With the vast development of Internet capacity and speed, as well as wide adoption of media technologies in people's daily life, a large amount of videos have been surging, and need to be efficiently processed or organized based on interest. The human visual perception system could, without difficulty, interpret and recognize thousands of events in videos, despite high level of video object clutters, different types of scene context, variability of motion scales, appearance changes, occlusions and object interactions. For a computer vision system, it has been very challenging to achieve automatic video event understanding for decades. Broadly speaking, those challenges include robust detection of events under motion clutters, event interpretation under complex scenes, multi-level semantic event inference, putting events in context and multiple cameras, event inference from object interactions, etc.

In recent years, steady progress has been made towards better models for video event categorisation and recognition, e.g., from modelling events with bag of spatial temporal features to discovering event context, from detecting events using a single camera to inferring events through a distributed camera network, and from low-level event feature extraction and description to high-level semantic event classification and recognition. Nowadays, text based video retrieval is widely used by commercial search engines. However, it is still very difficult to retrieve or categorise a specific video segment based on their content in a real multimedia system or in surveillance applications. To advance the progress further, we must adapt recent or existing approaches to find new solutions for intelligent video understanding.

This book aims to present state-of-the-art research advances of video event understanding technologies. It will provide researchers and practitioners a rich resource for future research directions and successful practice. It could also serve as a reference tool and handbook for researchers in a number of applications including visual surveillance, human-computer interaction, and video search and indexing etc. Its potential audience will be composed of active researchers and practitioners as well as graduate students working on video analysis in various disciplines such as computer vision, pattern recognition, information security, artificial intelligence, etc.

In Chapter 1, Vatavu addresses a double view of understanding meaningful events in gesture based interaction: events that specify gestures together with intelligent algorithms that detect them in video sequences; gestures, that once recognized and accordingly interpreted by the system, become important events in the human-computer dialogue specifying the common understanding that was

established. The chapter follows the duality aspect of events from the system as well as the human perspective contributing to the present understanding of gestures in human-computer interaction.

In Chapter 2, Yu and Zhang present a motion segmentation approach based on the subspace segmentation technique, the generalized PCA. By incorporating the cues from the neighbourhood of intensity edges of images, motion segmentation is solved under an algebra framework. They propose an effective post-processing procedure, which can detect the boundaries of motion layers and further determine the layer ordering.

In Chapter 3, Zhou presents a strategy based on human gait to achieve efficient tracking, recovery of ego-motion and 3-D reconstruction from an image sequence acquired by a single camera attached to a pedestrian. In the first phase, the parameters of the human gait are established by a classical frame-by-frame analysis, using a generalised least squares (GLS) technique. In the second phase, this gait model is employed within a "predict-correct" framework using a maximum a posterior, expectation maximization (MAP-EM) strategy to obtain robust estimates of the ego-motion and scene structure, while continuously refining the gait model.

In Chapter 4, Mattivi and Shao apply the Local Binary Pattern on Three Orthogonal Planes (LBP-TOP) descriptor to the field of human action recognition. They use LBP and CS-LBP techniques combined with gradient and Gabor images. Several modifications and extensions to the descriptor are further developed.

In Chapter 5, Zhuang et al. present an efficient object localization approach based on the Gaussianized vector representation following a branch-and-bound search scheme introduced by Lampert et al. In particular, they design a quality bound for rectangle sets characterized by the Gaussianized vector representation for fast hierarchical search. Further, they propose incorporating a normalization approach that suppresses the variation within the object class and the background class. This method outperforms previous work using the histogram-of-keywords representation for object localisation.

In Chapter 6, Zhang and Gong present a framework for robust people detection in highly cluttered scenes with low resolution image sequences. Their model utilises both human appearance and their long-term motion information. Preliminary studies demonstrate the method achieved good results under challenging conditions.

In Chapter 7, Hervieu and Bouthemey describe object-based approach for temporal analysis of sports videos using player's trajectories. An original hierarchical parallel semi-Markov model (HPaSMM) is proposed. Such probabilistic graphical models help taking into account low level temporal causalities of trajectories features as well as upper level temporal transitions between activity phases. It can be used for applications of sports video semantic-based understanding such that segmentation, summarization and indexing.

In Chapter 8, Davis et al. describe an experimental system for the recognition of human faces from surveillance video. Their system detects faces using the Viola-Jones face detector, and then extracts local features to build a shape-based

feature vector. Consideration was given to improving the performance and accuracy of both the detection and recognition steps.

In Chapter 9, Odashima et al. propose an object movement detection method in a household environment via the stable changes of images. To detect object placements and object removals robustly, the method employs the layered background model and the edge subtraction based classification method. In addition, to classify objects and non-objects robustly though the changed regions are occluded, the method uses motion history of the regions.

In Chapter 10, Ali et al. provide a survey on BBC Dirac Video Codec which can be use for compressing high resolution files, broadcasting, live video streaming, pod casting, and desktop production. This survey not only provides an in-deep description of different version of Dirac Video Codec but also explain the algorithmic explanation of Dirac at implementation level. It aims to help to new researchers who are working to understand BBC Dirac video codec but also provide them future directions and ideas to enhance features of BBC Dirac video codec.

Editors

Jianguo Zhang
Ling Shao
Lei Zhang
Graeme A. Jones

Contents

The Understanding of Meaningful Events in Gesture-Based Interaction	1
<i>Radu-Daniel Vatavu</i>	
Apply GPCA to Motion Segmentation	21
<i>Hongchuan Yu, Jian J. Zhang</i>	
Gait Analysis and Human Motion Tracking	39
<i>Huiyu Zhou</i>	
Spatio-temporal Dynamic Texture Descriptors for Human Motion Recognition	69
<i>Riccardo Mattivi, Ling Shao</i>	
Efficient Object Localization with Variation-Normalized Gaussianized Vectors	93
<i>Xiaodan Zhuang, Xi Zhou, Mark A. Hasegawa-Johnson, Thomas S. Huang</i>	
Fusion of Motion and Appearance for Robust People Detection in Cluttered Scenes	111
<i>Jianguo Zhang, Shaogang Gong</i>	
Understanding Sports Video Using Players Trajectories	125
<i>Alexandre Hervieu, Patrick Bouthemy</i>	
Real-Time Face Recognition from Surveillance Video	155
<i>Michael Davis, Stefan Popov, Cristina Surlea</i>	

Event Understanding of Human-Object Interaction: Object Movement Detection via Stable Changes	195
<i>Shigeyuki Odashima, Taketoshi Mori, Masamichi Simosaka, Hiroshi Noguchi, Tomomasa Sato</i>	

Survey of Dirac: A Wavelet Based Video Codec for Multiparty Video Conferencing and Broadcasting	211
<i>Ahtsham Ali, Nadeem A. Khan, Shahid Masud, Syed Farooq Ali</i>	

Erratum

Event Understanding of Human-Object Interaction: Object Movement Detection via Stable Changes	E1
<i>Shigeyuki Odashima, Taketoshi Mori, Masamichi Simosaka, Hiroshi Noguchi, Tomomasa Sato</i>	

Author Index	249
-------------------------------	------------