An Introduction to the Special Issue on Event Analysis in Videos

Shih-Fu Chang, *Fellow, IEEE*, Jiebo Luo, *Senior Member, IEEE*, Steve Maybank, Dan Schonfeld, and Dong Xu, *Member, IEEE*

I. Introduction

INTEREST from industry and academia has increased dramatically over recent years in the challenging area of event analysis and recognition from various video sources including sports, surveillance, user-generated video, etc. Video event analysis and recognition is a critical task in many applications such as detection of sporting highlights, incident detection in surveillance video, indexing, retrieval and summarization of video databases, and human-computer interaction. This special issue aims to capture the latest advances by the research community working in the area of video event analysis.

The call for papers was enthusiastically greeted by the research community and we received over seventy submissions. The special issue presents 16 articles which provide fundamental contributions in a wide range of topics in video event analysis: 1) human action and activity recognition; 2) motion trajectory analysis; 3) video content analysis and pattern mining; 4) audio-visual multi-modal analysis; and 5) video analysis applications. An overview of the organization and a brief summary of the articles selected for publication in the special issue are provided below.

II. Organization and Overview

A. Human Action and Activity Recognition

The invited paper by Turaga *et al.* presents a comprehensive survey, entitled: "Machine recognition of human activities: A survey," covering research efforts on human action and activity recognition over the past decade. In the area of human action recognition, the authors review and classify the prior work in terms of nonparametric, volumetric and parametric approaches. The presentation of the work in the area of human activity recognition is organized according to graphical model-based techniques, syntactic approaches, as well as knowledge and logic-based methods. In addition, other related issues, applications and potential future directions are also discussed.

The article by Zhou *et al.*, entitled: "Activity analysis, summarization, and visualization for indoor human activity monitoring," developed a prototype system for video-based elderly care monitoring. For this application, they proposed new techniques on multiple levels including silhouette extraction and human tracking techniques at the object level, an adaptive learning method at the feature level, a human action recognition scheme at the action level, as well as summarization and visualization techniques at the presentation level.

In the article "Expandable data-driven graphical modeling of human actions based on salient postures," Li *et al.* proposed a probabilistic, graphical model to encode actions in a weighted directed graph, referred to as an action graph. In an action graph, each node represents a salient posture modeled by a Gaussian mixture model (GMM). The authors also presented automatic methods to learn the action graph and the salient postures as well as effective techniques to recognize new actions using action graphs.

In "Combining fuzzy vector quantization with linear discriminant analysis for continuous human movement recognition," Nikolaos *et al.* proposed to combine fuzzy vector quantization (FVQ) and linear discriminant analysis (LDA) for continuous human movement recognition. The authors represented human movement as a sequence of dynemes, which are defined as the smallest unit of human motion. The authors also presented promising experimental results on the "Weizmann" dataset as well as a new dataset.

The article by Thome *et al.*, entitled: "A real-time, multiview fall detection system: A LHMM-based approach," explores the automatic detection of a falling person in a video sequence based on a multiview approach. The authors rely on a layered hidden Markov model (LHMM) to model motion and solve the inference problem. Independently processed data is used to detect, track, and extract features from each camera view. The processed data from multiple cameras is fused using a centralized unit for posture classification. Optimal camera placement allows for the best possible detection of falling persons in unknown environments.

B. Motion Trajectory Analysis

In the article "A statistical video content recognition method using an invariant feature on object trajectories," Hervieu *et al.* rely on invariant features of object trajectories for recognition of video content. Trajectories in video sequences are described using local features such as curvature and speed. The features are invariant under translation, rotation and scaling. The temporal sequences of feature values are modeled and compared using HMMs.

The article "Trajectory-based anomalous event detection" by Piciarelli *et al.* presents a method for anomalous event detection that relies on motion trajectory analysis to identify video events which differ from typical event patterns. They use a single-class

Digital Object Identifier 10.1109/TCSVT.2008.2007023

1051-8215/$25.00 © 2008 IEEE

support vector machine clustering to develop a method for identification of anomalous trajectories, especially when the number of outliers in the training data is unknown.

Anjum and Cavallaro present in "Multifeature object trajectory clustering for video analysis" a multifeature motion trajectory clustering algorithm for estimation of typical patterns and isolated outliers. The proposed algorithm relies on nonparametric clustering and information fusion to identify normal and abnormal event patterns. Clustering is performed by using the mean-shift algorithm to locate the modes of a feature-space representation of the motion trajectories. The determination of whether a motion trajectory is labeled as normal or abnormal is evaluated based on the density or sparsity of the trajectories among the clusters.

In the article "Event detection using trajectory clustering and 4-D histograms" by Jung *et al.*, a method for event detection based on trajectory clustering and 4-D histograms is proposed. The proposed method relies on global motion features to cluster motion trajectories. A histogram representation of the position and velocity of the tracked objects is associated with each cluster. The resulting histogram representation is used to ascertain the coherence of motion trajectories and previously tracked objects in the training data for detection of unusual motion patterns and video events.

C. Video Content Analysis and Pattern Mining

Zhou and Zhang, in the article entitled "An ICA mixture hidden Markov model for video content analysis," present a new theoretical framework based on the hidden Markov model (HMM) and independent component analysis (ICA) mixture model for content analysis of video sequences. A mixture of non-Gaussian components, each provided by an ICA mixture representation, is used in the proposed model and provides a superior representation of video components which captures their independence across video frames. The proposed model is subsequently used to derive maximum likelihood algorithms to detect and recognize video events such as recurrent patterns.

The article by Shen *et al.*, entitled: "Modality mixture projections for semantic video event detection," investigated modality mixture projections for semantic event detection, where multimodal information is expected to boost the performance. The authors use a subspace selection technique to achieve higher speed and accuracy. With the proposed modality mixture projections, feature vectors presenting different modalities associated with the video are projected onto a unified subspace, where the recognition takes place. The proposed framework was validated in comparisons to existing approaches using experimental results on both soccer video and TRECVID news video collections.

In the article "Mining recurrent events through forest growing," Yuan *et al.* propose a new approach to search a video sequence for recurring events, that is, subsequences which are similar to each other in appearance. The video sequence is reduced to a sequence of N video primitives. For each primitive, the K best matching primitives from the sequence are found. Recurring events are identified by searching the resulting $K \times N$ matrix for sequences of matching primitives, where adjacent primitives in each sequence appear in consecutive columns of the matrix.

D. Audio-Visual Multi-Modal Analysis

Vajaria *et al.* proposed in "Exploring co-occurrence between speech and body movement for audio-guided video localization" a method to localize speakers in meeting rooms recorded by using a single stationary camera and a single microphone. The authors utilized the long term co-occurrence of sounds and body motion for localization, mainly based on the observation that a talking person typically moves various parts of his body more than he does when sitting and listening. The experiments on a 21-h real video demonstrated that the proposed method outperforms the prior work.

In the article "Audio-assisted movie dialogue detection," Kotti *et al.* developed novel techniques using audio features and statistical models to detect dialogue events in movies. Actor indicator functions, pointing to the presence of an actor's speech at a specific time, and their correlations were used as inputs for classification of dialogue events.

E. Video Analysis Applications

The article by Han *et al.*, entitled: "Broadcast court-net sports video analysis using fast 3-D camera modeling," developed a framework for analyzing the court-net sports video content. They used a camera calibration process and 3-D modeling to map the 3-D real-world scene to the 2-D image domain, and then investigated the relations among players, playing-field, and occurrences of semantic events.

In the article "Semantic analysis for automatic event recognition and segmentation of wedding ceremony videos," Cheng *et al.* investigated a specific problem domain of wedding ceremony videos. The system developed automatically segments a wedding ceremony video into a sequence of recognizable wedding sub-events, e.g., the wedding kiss. Next, based on the domain knowledge of wedding customs, the system utilizes statistical models built upon a set of audiovisual features to classify thirteen wedding sub-events. These features are related to the wedding contexts of speech/music types, applause activities, picture-taking activities, and leading roles, while the models take into account both the fitness of observed features and the temporal rationality of event ordering.

III. Discussion

We believe that the articles selected for publication in this special issue provide an overview of the state-of-the-art in the field of video event analysis. We hope that this collection of papers will serve as a catalyst for future research efforts in the area of event analysis in video sequences.

Acknowledgment

We would like to thank all of the authors for their outstanding contributions and for sharing with us their interesting and innovative work on video event analysis. We are also grateful to the reviewers for their invaluable and insightful evaluations that helped shape this special issue. We extend our gratitude to Lauren Caruso from the editorial staff of TCSVT who has provided us with tremendous support and help. Finally, we are particularly grateful to the Editor-in-Chief, Prof. Chang Wen Chen, for his encouragement, guidance, and support throughout the entire process that has led to the publication of this special issue.

Shih-Fu Chang
Digital Video and Multimedia Laboratory
Columbia University
New York, NY 10027 USA

Jiebo Luo
Kodak Research Laboratories
Rochester, NY 14650 USA

Steve Maybank
Birkbeck College
University of London
London WC1E 7HX, U.K.

Dan Schonfeld
Department of Electrical and Computer Engineering
University of Illinois
Chicago, IL 60625 USA

Dong Xu
Nanyang Technical University
639798 Singapore

Shih-Fu Chang (M'93–F'04) leads the Digital Video and Multimedia Laboratory in the Department of Electrical Engineering of Columbia University, conducting research in multimedia content analysis, image/video search, multimedia forgery detection, and biomolecular image informatics. Systems developed by his group have been widely used, including VisualSEEk, WebSEEk for visual search, TrustFoto for online image authentication, and Columbia374 for large scale semantic visual concept detection. His group has received several best paper awards from the IEEE, ACM, and SPIE.

Dr. Chang is Editor-in-Chief of IEEE *Signal Processing Magazine* (2006–2008), a recipient of IEEE Kiyo Tomiyasu Award, Navy ONR Young Investigator Award, IBM Faculty Award, and a NSF CAREER Award.

Jiebo Luo (M'96–SM'99) received the B.S. degree from the University of Science and Technology of China, Heifei, China, in 1989 and the Ph.D. degree from the University of Rochester, Rochester, NY, in 1995, both in electrical engineering.

He is a Senior Principal Scientist with the Kodak Research Laboratories in Rochester, NY. His research interests include signal and image processing, machine learning, computer vision, multimedia data mining, and computational photography. He has authored over 130 technical papers and holds over 40 U.S. patents.

Dr. Luo has been involved in organizing numerous technical conferences sponsored by IEEE, ACM, and SPIE. Currently, he is on the Editorial Boards of the IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE, the IEEE TRANSACTIONS ON MULTIMEDIA, PATTERN RECOGNITION, and *Journal of Electronic Imaging*. He is a Kodak Distinguished Inventor, a winner of the 2004 Eastman Innovation Award, a member of ACM, a Senior Member of the IEEE, and a Fellow of SPIE.

Steve Maybank received the B.A. degree in mathematics (first class hons) from Cambridge University, Cambridge, U.K., in 1976, the diploma in computer science (with distinction) and the Ph.D. degree in computer science from Birkbeck College, University of London, London, U.K., in 1983 and 1988.

He was a GEC research scientist (1980–1995), Reading University lecturer (1995–2003) and from January 2004, professor at Birkbeck College, University of London. His research interests include computer vision, image processing, visual surveillance, statistics and information theory. He has published over 100 papers and one book.

Dan Schonfeld received the B.S. degree in electrical engineering and computer science from the University of California at Berkeley, and the M.S. and Ph.D. degrees in electrical and computer engineering from the Johns Hopkins University, Baltimore, MD, in 1986, 1988, and 1990, respectively.

In 1990, he joined the University of Illinois at Chicago, where he is currently a Professor in the Department of Electrical and Computer Engineering.

Dr. Schonfeld was co-author of papers that won the Best Student Paper Awards in SPIE VCIP 2006 and IEEE ICIP 2006 and 2007. He has served as Associate Editor of the IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS FOR VIDEO TECHNOLOGY, the IEEE TRANSACTIONS ON IMAGE PROCESSING, and the IEEE TRANSACTIONS ON SIGNAL PROCESSING. His current research interests are in multidimensional signal processing; image and video analysis; computer vision; and genomic signal processing.

Dong Xu (M'07) received the B.Eng. and Ph.D. degrees from the Electronic Engineering and Information Science Department, University of Science and Technology of China, Hiefei, China, in 2001 and 2005, respectively.

During his Ph.D. studies, he worked with Microsoft Research Asia and The Chinese University of Hong Kong. He also spent one year at Columbia University, New York, as a postdoctoral research scientist. His research interests include computer vision, pattern recognition, statistical learning and multimedia content analysis. He is currently an Assistant Professor at Nanyang Technological University, Singapore.