Dan Schonfeld, Caifeng Shan, Dacheng Tao, and Liang Wang (Eds.)

Video Search and Mining

Studies in Computational Intelligence, Volume 287

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. 266. Akitoshi Hanazawa, Tsutom Miki, and Keiichi Horio (Eds.)
Brain-Inspired Information Technology, 2009
ISBN 978-3-642-04024-5

Vol. 267. Ivan Zelinka, Sergej Celikovský, Hendrik Richter, and Guanrong Chen (Eds.)
Evolutionary Algorithms and Chaotic Systems, 2009
ISBN 978-3-642-10706-1

Vol. 268. Johann M.Ph. Schumann and Yan Liu (Eds.)
Applications of Neural Networks in High Assurance Systems, 2009
ISBN 978-3-642-10689-7

Vol. 269. Francisco Fernández de de Vega and Erick Cantú-Paz (Eds.)
Parallel and Distributed Computational Intelligence, 2009
ISBN 978-3-642-10674-3

Vol. 270. Zong Woo Geem
Recent Advances In Harmony Search Algorithm, 2009
ISBN 978-3-642-04316-1

Vol. 271. Janusz Kacprzyk, Frederick E. Petry, and Adnan Yazici (Eds.)
Uncertainty Approaches for Spatial Data Modeling and Processing, 2009
ISBN 978-3-642-10662-0

Vol. 272. Carlos A. Coello Coello, Clarisse Dhaenens, and Laetitia Jourdan (Eds.)
Advances in Multi-Objective Nature Inspired Computing, 2009
ISBN 978-3-642-11217-1

Vol. 273. Fatos Xhafa, Santi Caballé, Ajith Abraham, Thanasis Daradoumis, and Angel Alejandro Juan Perez (Eds.)
Computational Intelligence for Technology Enhanced Learning, 2010
ISBN 978-3-642-11223-2

Vol. 274. Zbigniew W. Raś and Alicja Wieczorkowska (Eds.)
Advances in Music Information Retrieval, 2010
ISBN 978-3-642-11673-5

Vol. 275. Dilip Kumar Pratihar and Lakhmi C. Jain (Eds.)
Intelligent Autonomous Systems, 2010
ISBN 978-3-642-11675-9

Vol. 276. Jacek Mańdziuk
Knowledge-Free and Learning-Based Methods in Intelligent Game Playing, 2010
ISBN 978-3-642-11677-3

Vol. 277. Filippo Spagnolo and Benedetto Di Paola (Eds.)
European and Chinese Cognitive Styles and their Impact on Teaching Mathematics, 2010
ISBN 978-3-642-11679-7

Vol. 278. Radomir S. Stankovic and Jaakko Astola
From Boolean Logic to Switching Circuits and Automata, 2010
ISBN 978-3-642-11681-0

Vol. 279. Manolis Wallace, Ioannis E. Anagnostopoulos, Phivos Mylonas, and Maria Bielikova (Eds.)
Semantics in Adaptive and Personalized Services, 2010
ISBN 978-3-642-11683-4

Vol. 280. Chang Wen Chen, Zhu Li, and Shiguo Lian (Eds.)
Intelligent Multimedia Communication: Techniques and Applications, 2010
ISBN 978-3-642-11685-8

Vol. 281. Robert Babuska and Frans C.A. Groen (Eds.)
Interactive Collaborative Information Systems, 2010
ISBN 978-3-642-11687-2

Vol. 282. Husrev Taha Sencar, Sergio Velastin, Nikolaos Nikolaidis, and Shiguo Lian (Eds.)
Intelligent Multimedia Analysis for Security Applications, 2010
ISBN 978-3-642-11754-1

Vol. 283. Ngoc Thanh Nguyen, Radoslaw Katarzyniak, and Shyi-Ming Chen (Eds.)
Advances in Intelligent Information and Database Systems, 2010
ISBN 978-3-642-12089-3

Vol. 284. Juan R. González, David Alejandro Pelta, Carlos Cruz, Germán Terrazas, and Natalio Krasnogor (Eds.)
Nature Inspired Cooperative Strategies for Optimization (NICSO 2010), 2010
ISBN 978-3-642-12537-9

Vol. 285. Roberto Cipolla, Sebastiano Battiato, and Giovanni Maria Farinella (Eds.)
Computer Vision, 2010
ISBN 978-3-642-12847-9

Vol. 286. Zeev Volkovich, Alexander Bolshoy, Valery Kirzhner, and Zeev Barzily
Genome Clustering, 2010
ISBN 978-3-642-12951-3

Vol. 287. Dan Schonfeld, Caifeng Shan, Dacheng Tao, and Liang Wang (Eds.)
Video Search and Mining, 2010
ISBN 978-3-642-12899-8

Dan Schonfeld, Caifeng Shan, Dacheng Tao,
and Liang Wang (Eds.)

Video Search and Mining

Springer

Dan Schonfeld
Multimedia Communications Laboratory
Department of Electrical & Computer Engineering
University of Illinois at Chicago
Room 1020 SEO (M/C 154)
851 South Morgan Street
Chicago, IL 60607-7053
USA
E-mail: dans@uic.edu

Caifeng Shan
Philips Research
High-Tech Campus 36
5656 AE, Eindhoven
The Netherlands
E-mail: caifeng.shan@philips.com

Dacheng Tao
Department of Computing
Hong Kong Polytechnic University
PQ704, 7/F, Building P
Hung Hom, Kowloon, Hong Kong
China
E-mail: dacheng.tao@gmail.com

Liang Wang
Department of Computer Science
University of Bath, BA2 7AY
United Kingdom
Email: lw356@cs.bath.ac.uk

ISBN 978-3-642-12899-8 e-ISBN 978-3-642-12900-1

DOI 10.1007/978-3-642-12900-1

Studies in Computational Intelligence ISSN 1860-949X

Library of Congress Control Number: Applied for

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

As cameras become more pervasive in our daily life, vast amounts of video data are generated. The popularity of YouTube and similar websites such as Tudou and Youku provides strong evidence for the increasing role of video in society. One of the main challenges confronting us in the era of information technology is to effectively rely on the huge and rapidly growing video data accumulating in large multimedia archives. Innovative video processing and analysis techniques will play an increasingly important role in resolving the difficult task of video search and retrieval. A wide range of video-based applications have benefited from advances in video search and mining including multimedia information management, human-computer interaction, security and surveillance, copyright protection, and personal entertainment, to name a few.

This book provides an overview of emerging new approaches to video search and mining based on promising methods being developed in the computer vision and image analysis community. Video search and mining is a rapidly evolving discipline whose aim is to capture interesting patterns in video data. It has become one of the core areas in the data mining research community. In comparison to other types of data mining (e.g. text), video mining is still in its infancy. Many challenging research problems are facing video mining researchers. For example, how to extract knowledge from spatio-temporal data? How to infer high-level semantic concepts from low-level features in videos? How to exploit unlabeled and untagged video data? The use of classical data mining techniques for video data is impractical due to the massive volume of high-dimensional video data. To address these difficult challenges, it is necessary to develop search and mining techniques and methods that are suitable for video data.

The objective of this book is to present the latest advances in video search and mining covering both theoretical approaches and practical applications. The book provides researchers and practitioners a comprehensive understanding of the start-of-the-art in video search and mining techniques and a resource for potential applications and successful practice. This book can also serve as an important reference tool and handbook for researchers and practitioners in video search and mining.

The target audience of this book is mainly engineers and students working on video analysis in various disciplines, e.g. computer vision, pattern recognition, information technology, image processing, and artificial intelligence. The book is intended to be accessible to a broader audience including researchers and practicing professionals working in video applications such as video surveillance, video retrieval, etc.

The origin of this book stems from the immense success of the First International Workshop on Video Mining (VM'08), held in conjunction with the IEEE International Conference on Data Mining 2008. This workshop gathered experts from different fields working on video search and mining.

Organization and Themes

The book comprises both theoretical advances and applications in video search and mining. The organization of the book reflects the combination of analytical and practical topics addressed throughout the book. We have divided the book chapters into five parts; each addresses a specific theme in video search and mining. The five themes presented include motion trajectory analysis, high-dimensional video representation, semantic video analysis, personalized video, and video mining.

Part I. Motion Trajectory Analysis: Object motion trajectories describe the rich dynamic content of video data. Motion trajectory analysis plays an important role in video mining and has been exploited for various applications, e.g. video retrieval, video summarization, video surveillance, traffic monitoring, and sports analysis.

Chapter 1 is focused on the latest research in motion trajectory analysis for video search and mining. The main methodologies for the description of motion trajectories, as well as the indexing techniques and similarity metrics used in the retrieval process are introduced and examined through a comparative analysis, application examples, and discussion on future trends. In Chapter 2, trajectory clustering methods for learning activity models are introduced with a focus on parametric and non-parametric partition algorithms. A soft partition algorithm based on non-parametric mean-shift clustering is proposed and validated. The use of one- and multi-dimensional hidden Markov models (HMMs) for video classification and recognition is also introduced.

Several aspects of motion trajectory analysis including their representation, indexing, similarity, invariance, and application are addressed in Chapter 3. For instance, a representation of motion trajectories that is invariant to camera view is introduced based on null-space invariants. The representation of motion trajectories includes both isolated motion trajectories describing the dynamics of a single object as well as multiple motion trajectories characterizing the interaction among a group of objects in complex events. Several methods for efficient indexing based on matrix and tensor decomposition and various similarity measures for efficient storage, retrieval and classification of motion trajectories are reviewed.

Examples of the use of motion trajectory analysis in real world applications provide a vivid illustration and help the reader gain a deeper insight into the potential of motion trajectory analysis in video search and mining.

Part II. High-Dimensional Video Representation: Because of the innate abundance of information in video sequences, efficient representation of video data has long been a subject of intense interest in video analysis. Video representations are invariably high-dimensional, and thus a great deal of effort is required for processing video data. The representation of video in high-dimensional spaces is the topic of this part of the book.

Chapter 4 explores extraction of features that reflect three-dimensional structural information embedded in videos. Three-dimensional features venture beyond traditional video features, e.g. *two-dimensional appearance-based* features or spatio-temporal features. These three-dimensional features expand the kind of information captured from videos, which is essential in tasks such as video mining and retrieval. The main difficulty posed by this approach is the need to infer accurate three-dimensional information. This limitation is due to the fact that typical video footage does not convey prior knowledge about the scene configuration or camera calibration. Several recent methods that address the challenge posed by three-dimensional feature-based video analysis are reviewed including simultaneous localization and mapping, structure-from-motion, and 3D reconstruction. Illustrative use of tensor-based representation in practical applications including shot boundary detection, object recognition, content-based video retrieval as well as human activity recognition are presented, and limitations of the state-of-the-art techniques are discussed.

Models designed to understand, organize and utilize acquired high-dimensional video representation are considered. The goal is to extract semantically meaningful information from a large collection of samples in a high-dimensional space. Specifically, in video analysis, one of the fundamental aims is to recognize actions or persons. Statistical inference has been very successful in performing these tasks. Extension of statistical inference tools to manifolds in high-dimensional spaces is discussed in Chapter 5. Manifold is critical to the representation of video data, which is distributed over a small fraction of the high-dimensional space. Once the distribution of video data on manifolds has been captured, efficient statistical inference and learning methods can be developed. Empirical evaluation of various manifold-based video analysis techniques for applications such as video understanding is presented.

Part III. Semantic Video Analysis: The topic of understanding videos by semantic learning has been an important research area over the past decade. The formalism of semantic activity analysis including ontological representation of domain knowledge, statistical models and logical languages to describe activities, and the integration of semantic schemes will be discussed in Chapter 6. The focus is on techniques for addressing challenges in automatic activity recognition and abnormality detection such as robust primitive action detection, multi-agent interaction, higher-/lower-level action mapping, etc.

Video genre inference, a technique that automatically clusters and tags video data and thus facilitates user browsing, search and retrieval, provides an important practical illustration of semantic inference in video analysis. A comprehensive introduction of the necessary background for video genre inference as well as recent developments in the use of task-specific features, reflecting the video capture and camera configuration, is presented in Chapter 7.

Chapter 8 explores the problem of semantic visual learning from weakly labeled web video. A general tenet of learning theory is that an infallible supervisor always provides useful information that yields improved results. In the context of learning from videos, however, it is generally too expensive, and often not possible, to attain complete and accurate information by a supervisor. A technique for robust learning from weakly tagged video data (as available from the web) is described. The interesting feature of the technique illustrated is the mechanism used for automatic identification and filtering of irrelevant information. Another important component of the method described is the use of active learning to allow users to intervene in order to further improve the performance at a controllable cost.

Part IV. Personalized Video: Personalization is an increasingly important trait in recent multimedia applications. The theme in this part of the book will investigate the role of personalization in video analysis. In particular, the use of video-based face recognition for personalized video as well as a personalized recommendation system for broadcast news will be described.

Automatic face recognition is one of the most active research areas in computer vision. It allows automatic identification and verification of a person from a static image or video sequences. In contrast to traditional static image-based approaches, video-based face recognition technology utilizes the abundant video information, leading to more accurate and robust face recognition methods, and thus provides an invaluable tool for personalized video applications. A comprehensive survey of video-based face recognition and retrieval techniques for personalized video applications is presented in Chapter 9.

Because of the large number of broadcast channels and TV news programs, finding news videos of interest can be a difficult task. Chapter 10 presents an interactive framework for personalized news video recommendation. The personalized system allows news seekers to access programs of interest from large-scale news video archives more effectively. In this framework, multiple media sources (i.e. audio, video, and closed captions) are integrated to capture news topics, and their inter-topic contextual relationships are visualized to enable news seekers to interactively find news topics of interest.

Part V. Video Mining: The final theme of the book is focused on empirical tasks in video mining. The scope of the discussion ranges from low-level tasks, such as motion detection and re-occurrence identification to semantic-level tasks.

Motion is an important feature in content-based video parsing for high-level understanding. Mining motion information from video data is a critical task in video analysis. In Chapter 11, a case study is illustrated to introduce a unique technique of independent motion segmentation. A holistic, in-compression approach is presented,

thus attaining high-efficiency while providing very good performance in mining motion information from video data.

Another critical facet of video data is represented by pattern re-occurrence in video sequences. The problem of detection of (almost)-repetitive pattern occurrence in video footage is an important task in video mining. Examples of recurring video sequences include commercials, channel advertisements, channel intros, and newscast intros. Video occurrence identification requires integration of various fundamental video analysis techniques, including feature extraction, classifier training, and efficient search. Several technologies for recognizing predefined and unknown recurring video sequences are discussed for mining general videos and broadcast television in Chapters 12 and 13, respectively. A real-time recurring video sequence recognition system is also presented in Chapter 13.

Automatic video annotation is an important tool for video indexing and search. It has been well studied for several years. However, most existing efforts in video annotation have been conducted on small video corpuses and the size of concept vocabulary has been limited to tens or hundreds. The exponential increase in video data on the Web presents a great potential as well as new challenges for video annotation. Chapter 14 introduces the use of video mining techniques for automatic web-scale video annotation as well as annotating videos using large-sized natural language keywords. State-of-the-art techniques for video analysis, feature extraction and classification are presented. The focus is on the use of suitable techniques, key algorithms, and data structures, for extremely large-scale automatic video annotation. A new web-scale automatic video annotation method is presented in which the keywords are augmented based on previous annotations, by leveraging the large collection of video data. The effectiveness and efficiency of the method described are validated using experiments conducted on a corpus of over 1.2 million videos crawled from YouTube.

Acknowledgements

The preparation of this book has been an arduous task and required the devotion of many people. We would like to thank all the authors for their tremendous effort and dedication in preparing the book chapters. We would also like to thank all of the reviewers as well as the members of the editorial board of the Video Mining Workshop at ICDM'08. The invaluable comments and suggestions provided by the reviewers have helped shape the final book. Finally, we would like to thank Thomas Ditzinger, Senior Editor at Springer-Verlag Heidelberg, for his constant dedication and support throughout the preparation of this book.

Editors

Prof. Dan Schonfeld
Department of Electrical & Computer Engineering
University of Illinois at Chicago, Chicago, IL 60607-7053, USA
Email: dans@uic.edu
URL: http://www.ece.uic.edu/~ds

Dr. Caifeng Shan
Philips Research
High-Tech Campus 36, 5656 AE, Eindhoven, The Netherlands
Email: caifeng.shan@philips.com
URL: http://www.dcs.qmul.ac.uk/~cfshan

Dr. Dacheng Tao
School of Computer Engineering
Nanyang Technological University, Singapore 639798
Email: dacheng.tao@ieee.org
URL: http://www.ntu.edu.sg/home/dctao

Dr. Liang Wang
Department of Computer Science
University of Bath, BA2 7AY, United Kingdom
Email: lw356@cs.bath.ac.uk
URL: http://www.cs.bath.ac.uk/~lw356/

Contents