Guoping Qiu
Clement Leung
Xiangyang Xue
Robert Laurini (Eds.)

LNCS 4781

Advances in Visual Information Systems

9th International Conference, VISUAL 2007
Shanghai, China, June 2007
Revised Selected Papers

Springer

Guoping Qiu Clement Leung
Xiangyang Xue Robert Laurini (Eds.)

Advances in Visual Information Systems

9th International Conference, VISUAL 2007
Shanghai, China, June 28-29, 2007
Revised Selected Papers

Springer

Volume Editors

Guoping Qiu
University of Nottingham
School of Computer Science
Jubilee Campus, Nottingham NG8 1BB, UK
E-mail: qiu@cs.nott.ac.uk

Clement Leung
Hong Kong Baptist University
Department of Computer Science
Kowloon Tong, Hong Kong, China
E-mail: clement@comp.hkbu.edu.hk

Xiangyang Xue
Fudan University
Department of Computer Science & Engineering
Shanghai Key Laboratory of Intelligent Information Processing
Shanghai 200433, China
E-mail: xyxue@fudan.edu.cn

Robert Laurini
Institut National des Sciences Appliquées (INSA) de Lyon
Laboratoire d'InfoRmatique en Image et Systèmes d'information (LIRIS)
Bât. Blaise Pascal, 7 av. Jean Capelle, 69621 Villeurbanne, France
E-mail: Robert.Laurini@insa-lyon.fr

Library of Congress Control Number: 2007940861

CR Subject Classification (1998): I.4, I.5, I.2.6-10, I.3, H.3, H.5, H.2

LNCS Sublibrary: SL 6 – Image Processing, Computer Vision, Pattern Recognition, and Graphics

ISSN 0302-9743
ISBN-10 3-540-76413-5 Springer Berlin Heidelberg New York
ISBN-13 978-3-540-76413-7 Springer Berlin Heidelberg New York

Springer is a part of Springer Science+Business Media

springer.com

Printed in Germany

Typesetting: Camera-ready by author, data conversion by Scientific Publishing Services, Chennai, India
Printed on acid-free paper SPIN: 12177349 06/3180 5 4 3 2 1 0

Lecture Notes in Computer Science 4781

Lecture Notes in Computer Science

Sublibrary 6: Image Processing, Computer Vision, Pattern Recognition, and Graphics

Vol. 4844: Y. Yagi, S.B. Kang, I.S. Kweon, H. Zha (Eds.), Computer Vision – ACCV 2007, Part II. XXVIII, 915 pages. 2007.

Vol. 4843: Y. Yagi, S.B. Kang, I.S. Kweon, H. Zha (Eds.), Computer Vision – ACCV 2007, Part I. XXVIII, 969 pages. 2007.

Vol. 4842: G. Bebis, R. Boyle, B. Parvin, D. Koracin, N. Paragios, S.-M. Tanveer, T. Ju, Z. Liu, S. Coquillart, C. Cruz-Neira, T. Müller, T. Malzbender (Eds.), Advances in Visual Computing, Part II. XXXIII, 827 pages. 2007.

Vol. 4841: G. Bebis, R. Boyle, B. Parvin, D. Koracin, N. Paragios, S.-M. Tanveer, T. Ju, Z. Liu, S. Coquillart, C. Cruz-Neira, T. Müller, T. Malzbender (Eds.), Advances in Visual Computing, Part I. XXXIII, 831 pages. 2007.

Vol. 4815: A. Ghosh, R.K. De, S.K. Pal (Eds.), Pattern Recognition and Machine Intelligence. XIX, 677 pages. 2007.

Vol. 4814: A. Elgammal, B. Rosenhahn, R. Klette (Eds.), Human Motion – Understanding, Modeling, Capture and Animation. X, 329 pages. 2007.

Vol. 4792: N. Ayache, S. Ourselin, A. Maeder (Eds.), Medical Image Computing and Computer-Assisted Intervention – MICCAI 2007, Part II. XLVI, 988 pages. 2007.

Vol. 4791: N. Ayache, S. Ourselin, A. Maeder (Eds.), Medical Image Computing and Computer-Assisted Intervention – MICCAI 2007, Part I. XLVI, 1012 pages. 2007.

Vol. 4781: G. Qiu, C. Leung, X. Xue, R. Laurini (Eds.), Advances in Visual Information Systems. XIII, 582 pages. 2007.

Vol. 4778: S.K. Zhou, W. Zhao, X. Tang, S. Gong (Eds.), Analysis and Modeling of Faces and Gestures. X, 305 pages. 2007.

Vol. 4756: L. Rueda, D. Mery, J. Kittler (Eds.), Progress in Pattern Recognition, Image Analysis and Applications. XXI, 989 pages. 2007.

Vol. 4738: A. Paiva, R. Prada, R.W. Picard (Eds.), Affective Computing and Intelligent Interaction. XVIII, 781 pages. 2007.

Vol. 4729: F. Mele, G. Ramella, S. Santillo, F. Ventriglia (Eds.), Advances in Brain, Vision, and Artificial Intelligence. XVI, 618 pages. 2007.

Vol. 4713: F.A. Hamprecht, C. Schnörr, B. Jähne (Eds.), Pattern Recognition. XIII, 560 pages. 2007.

Vol. 4679: A.L. Yuille, S.-C. Zhu, D. Cremers, Y. Wang (Eds.), Energy Minimization Methods in Computer Vision and Pattern Recognition. XII, 494 pages. 2007.

Vol. 4678: J. Blanc-Talon, W. Philips, D. Popescu, P. Scheunders (Eds.), Advanced Concepts for Intelligent Vision Systems. XXIII, 1100 pages. 2007.

Vol. 4673: W.G. Kropatsch, M. Kampel, A. Hanbury (Eds.), Computer Analysis of Images and Patterns. XX, 1006 pages. 2007.

Vol. 4642: S.-W. Lee, S.Z. Li (Eds.), Advances in Biometrics. XX, 1216 pages. 2007.

Vol. 4633: M. Kamel, A. Campilho (Eds.), Image Analysis and Recognition. XII, 1312 pages. 2007.

Vol. 4584: N. Karssemeijer, B. Lelieveldt (Eds.), Information Processing in Medical Imaging. XX, 777 pages. 2007.

Vol. 4569: A. Butz, B. Fisher, A. Krüger, P. Olivier, S. Owada (Eds.), Smart Graphics. IX, 237 pages. 2007.

Vol. 4538: F. Escolano, M. Vento (Eds.), Graph-Based Representations in Pattern Recognition. XII, 416 pages. 2007.

Vol. 4522: B.K. Ersbøll, K.S. Pedersen (Eds.), Image Analysis. XVIII, 989 pages. 2007.

Vol. 4485: F. Sgallari, A. Murli, N. Paragios (Eds.), Scale Space and Variational Methods in Computer Vision. XV, 931 pages. 2007.

Vol. 4478: J. Martí, J.M. Benedí, A.M. Mendonça, J. Serrat (Eds.), Pattern Recognition and Image Analysis, Part II. XXVII, 657 pages. 2007.

Vol. 4477: J. Martí, J.M. Benedí, A.M. Mendonça, J. Serrat (Eds.), Pattern Recognition and Image Analysis, Part I. XXVII, 625 pages. 2007.

Vol. 4472: M. Haindl, J. Kittler, F. Roli (Eds.), Multiple Classifier Systems. XI, 524 pages. 2007.

Vol. 4466: F.B. Sachse, G. Seemann (Eds.), Functional Imaging and Modeling of the Heart. XV, 486 pages. 2007.

Vol. 4418: A. Gagalowicz, W. Philips (Eds.), Computer Vision/Computer Graphics Collaboration Techniques. XV, 620 pages. 2007.

Vol. 4417: A. Kerren, A. Ebert, J. Meyer (Eds.), Human-Centered Visualization Environments. XIX, 403 pages. 2007.

Vol. 4391: Y. Stylianou, M. Faundez-Zanuy, A. Esposito (Eds.), Progress in Nonlinear Speech Processing. XII, 269 pages. 2007.

Vol. 4370: P.P. Lévy, B. Le Grand, F. Poulet, M. Soto, L. Darago, L. Toubiana, J.-F. Vibert (Eds.), Pixelization Paradigm. XV, 279 pages. 2007.

Vol. 4358: R. Vidal, A. Heyden, Y. Ma (Eds.), Dynamical Vision. IX, 329 pages. 2007.

Vol. 4338: P.K. Kalra, S. Peleg (Eds.), Computer Vision, Graphics and Image Processing. XV, 965 pages. 2006.

Vol. 4319: L.-W. Chang, W.-N. Lie (Eds.), Advances in Image and Video Technology. XXVI, 1347 pages. 2006.

Vol. 4292: G. Bebis, R. Boyle, B. Parvin, D. Koracin, P. Remagnino, A. Nefian, G. Meenakshisundaram, V. Pascucci, J. Zara, J. Molineros, H. Theisel, T. Malzbender (Eds.), Advances in Visual Computing, Part II. XXXII, 906 pages. 2006.

Vol. 4291: G. Bebis, R. Boyle, B. Parvin, D. Koracin, P. Remagnino, A. Nefian, G. Meenakshisundaram, V. Pascucci, J. Zara, J. Molineros, H. Theisel, T. Malzbender (Eds.), Advances in Visual Computing, Part I. XXXI, 916 pages. 2006.

Vol. 4245: A. Kuba, L.G. Nyúl, K. Palágyi (Eds.), Discrete Geometry for Computer Imagery. XIII, 688 pages. 2006.

Vol. 4241: R.R. Beichel, M. Sonka (Eds.), Computer Vision Approaches to Medical Image Analysis. XI, 262 pages. 2006.

Vol. 4225: J.F. Martínez-Trinidad, J.A. Carrasco Ochoa, J. Kittler (Eds.), Progress in Pattern Recognition, Image Analysis and Applications. XIX, 995 pages. 2006.

Vol. 4191: R. Larsen, M. Nielsen, J. Sporring (Eds.), Medical Image Computing and Computer-Assisted Intervention – MICCAI 2006, Part II. XXXVIII, 981 pages. 2006.

Vol. 4190: R. Larsen, M. Nielsen, J. Sporring (Eds.), Medical Image Computing and Computer-Assisted Intervention – MICCAI 2006, Part I. XXXVVIII, 949 pages. 2006.

Vol. 4179: J. Blanc-Talon, W. Philips, D. Popescu, P. Scheunders (Eds.), Advanced Concepts for Intelligent Vision Systems. XXIV, 1224 pages. 2006.

Vol. 4174: K. Franke, K.-R. Müller, B. Nickolay, R. Schäfer (Eds.), Pattern Recognition. XX, 773 pages. 2006.

Vol. 4170: J. Ponce, M. Hebert, C. Schmid, A. Zisserman (Eds.), Toward Category-Level Object Recognition. XI, 618 pages. 2006.

Vol. 4153: N. Zheng, X. Jiang, X. Lan (Eds.), Advances in Machine Vision, Image Processing, and Pattern Analysis. XIII, 506 pages. 2006.

Vol. 4142: A. Campilho, M. Kamel (Eds.), Image Analysis and Recognition, Part II. XXVII, 923 pages. 2006.

Vol. 4141: A. Campilho, M. Kamel (Eds.), Image Analysis and Recognition, Part I. XXVIII, 939 pages. 2006.

Vol. 4122: R. Stiefelhagen, J.S. Garofolo (Eds.), Multimodal Technologies for Perception of Humans. XII, 360 pages. 2007.

Vol. 4109: D.-Y. Yeung, J.T. Kwok, A. Fred, F. Roli, D. de Ridder (Eds.), Structural, Syntactic, and Statistical Pattern Recognition. XXI, 939 pages. 2006.

Vol. 4091: G.-Z. Yang, T. Jiang, D. Shen, L. Gu, J. Yang (Eds.), Medical Imaging and Augmented Reality. XIII, 399 pages. 2006.

Vol. 4073: A. Butz, B. Fisher, A. Krüger, P. Olivier (Eds.), Smart Graphics. XI, 263 pages. 2006.

Vol. 4069: F.J. Perales, R.B. Fisher (Eds.), Articulated Motion and Deformable Objects. XV, 526 pages. 2006.

Vol. 4057: J.P.W. Pluim, B. Likar, F.A. Gerritsen (Eds.), Biomedical Image Registration. XII, 324 pages. 2006.

Vol. 4046: S.M. Astley, M. Brady, C. Rose, R. Zwiggelaar (Eds.), Digital Mammography. XVI, 654 pages. 2006.

Vol. 4040: R. Reulke, U. Eckardt, B. Flach, U. Knauer, K. Polthier (Eds.), Combinatorial Image Analysis. XII, 482 pages. 2006.

Vol. 4035: T. Nishita, Q. Peng, H.-P. Seidel (Eds.), Advances in Computer Graphics. XX, 771 pages. 2006.

Vol. 3979: T.S. Huang, N. Sebe, M. Lew, V. Pavlović, M. Kölsch, A. Galata, B. Kisačanin (Eds.), Computer Vision in Human-Computer Interaction. XII, 121 pages. 2006.

Vol. 3954: A. Leonardis, H. Bischof, A. Pinz (Eds.), Computer Vision – ECCV 2006, Part IV. XVII, 613 pages. 2006.

Vol. 3953: A. Leonardis, H. Bischof, A. Pinz (Eds.), Computer Vision – ECCV 2006, Part III. XVII, 649 pages. 2006.

Vol. 3952: A. Leonardis, H. Bischof, A. Pinz (Eds.), Computer Vision – ECCV 2006, Part II. XVII, 661 pages. 2006.

Vol. 3951: A. Leonardis, H. Bischof, A. Pinz (Eds.), Computer Vision – ECCV 2006, Part I. XXXV, 639 pages. 2006.

Vol. 3948: H.I. Christensen, H.-H. Nagel (Eds.), Cognitive Vision Systems. VIII, 367 pages. 2006.

Vol. 3926: W. Liu, J. Lladós (Eds.), Graphics Recognition. XII, 428 pages. 2006.

Vol. 3872: H. Bunke, A.L. Spitz (Eds.), Document Analysis Systems VII. XIII, 630 pages. 2006.

Vol. 3852: P.J. Narayanan, S.K. Nayar, H.-Y. Shum (Eds.), Computer Vision – ACCV 2006, Part II. XXXI, 977 pages. 2006.

Vol. 3851: P.J. Narayanan, S.K. Nayar, H.-Y. Shum (Eds.), Computer Vision – ACCV 2006, Part I. XXXI, 973 pages. 2006.

Vol. 3832: D. Zhang, A.K. Jain (Eds.), Advances in Biometrics. XX, 796 pages. 2005.

Vol. 3736: S. Bres, R. Laurini (Eds.), Visual Information and Information Systems. XI, 291 pages. 2006.

Vol. 3667: W.J. MacLean (Ed.), Spatial Coherence for Visual Motion Analysis. IX, 141 pages. 2006.

Vol. 3417: B. Jähne, R. Mester, E. Barth, H. Scharr (Eds.), Complex Motion. X, 235 pages. 2007.

Vol. 2396: T.M. Caelli, A. Amin, R.P.W. Duin, M.S. Kamel, D. de Ridder (Eds.), Structural, Syntactic, and Statistical Pattern Recognition. XVI, 863 pages. 2002.

Vol. 1679: C. Taylor, A. Colchester (Eds.), Medical Image Computing and Computer-Assisted Intervention – MICCAI'99. XXI, 1240 pages. 1999.

Preface

The Visual Information Systems International Conference series is designed to provide a forum for researchers and practitioners from diverse areas of computing including computer vision, databases, human–computer interaction, information security, image processing, information visualization and mining, as well as knowledge and information management to exchange ideas, discuss challenges, present their latest results and to advance research and development in the construction and application of visual information systems. Following previous conferences held in Melbourne (1996), San Diego (1997), Amsterdam (1999), Lyon (2000), Taiwan (2002), Miami (2003), San Francisco (2004) and Amsterdam (2005), the Ninth International Conference on Visual Information Systems, VISUAL2007, was held in Shanghai, China, June 28–29, 2007.

Over the years, the visual information systems paradigm continues to evolve, and the unrelenting exponential growth in the amount of digital visual data underlines the escalating importance of how such data are effectively managed and deployed. VISUAL2007 received 117 submissions from 15 countries and regions. Submitted full papers were reviewed by more than 60 international experts in the field. This volume collects 54 selected papers presented at VISUAL2007. Topics covered in these papers include image and video retrieval, visual biometrics, intelligent visual information processing, visual data mining, ubiquitous and mobile visual information systems, visual semantics, 2D/3D graphical visual data retrieval and applications of visual information systems.

Two distinguished researchers delivered keynote talks at VISUAL2007. Wei-Ying Ma from Microsoft Research Asia gave a talk on "The Challenges and Opportunities of Mining Billions of Web Images for Search and Advertising." Michael Lew from Linden University, The Netherlands, gave a talk on "Visual Information Retrieval: Grand Challenges and Future Directions."

We would like to thank Hong Lu, Yue-Fei Guo and their team for the significant organization effort that they put in. We are grateful to the Department of Computer Science and Engineering, Fudan University for hosting the conference. In particular, we would like to express our gratitude to members of the Program Committee for their part in reviewing the papers to ensure a high-quality and successful conference.

September 2007

Guoping Qiu
Clement Leung
Xiang-Yang Xue
Robert Laurini

Organization

General Chairs

Robert Laurini
INSA of Lyon, France

Xiang-Yang Xue
Fudan University, China

Technical Program Chairs

Clement Leung
Hong Kong Baptist University, Hong Kong

Guoping Qiu
University of Nottingham, UK

Program Committee

Ching-chih Chen, Simmons College, USA
Yixin Chen, University of Mississippi, USA
Mingmin Chi, Fudan University, China
Zheru Chi, Hong Kong Polytechnic University, Hong Kong
Arjen P. de Vries, CWI, The Netherlands
Martin Dzbor, Knowledge Media Institute, The Open University, UK
Peter Enser, University of Brighton, UK
Jianping Fan, University of North Carolina at Charlotte, USA
Graham Finlayson, University of East Anglia, UK
Xiaodong Gu, Thomson R&D Centre, Beijing, China
Alan Hanjalic, Delft University of Technology, The Netherlands
Xian-Sheng Hua, Microsoft Research Asia, Beijing, China
Jesse Jin, Newcastle University, Australia
Joemon Jose, University of Glasgow, UK
Irwin King, The Chinese University of Hong Kong, Hong Kong
Markus Koskela, Dublin City University, Ireland
Igor V. Kozintsev, Intel Microprocessor Research Lab, USA
Jorma Laaksonen, Helsinki University of Technology, Finland
Kenneth Lam, Hong Kong Polytechnic University, Hong Kong
Robert Laurini, INSA, Lyon, France
Bongshin Lee, Microsoft Research, Redmond, USA
Wee-Kheng Leow, National University of Singapore, Singapore

Clement Leung, Victoria University, Melbourne, Australia
Michael Lew, University of Leiden, The Netherlands
Ze-Nian Li, Simon Fraser University, Canada
Rainer W. Lienhart, University of Mannheim, Germany
Tie-Yan Liu, Microsoft Research Asia, China
Hong Lu, Fudan University, China
Stephane Marchand-Maillet, University of Geneva, Switzerland
Graham Martin, Warwick University, UK
Jean Martinet, National Institute of Informatics, Japan
Dalibor Mitrovic, Vienna University of Technology, Austria
Keith Nesbitt, Charles Stuart University, Australia
Chong-Wah Ngo, City University of Hong Kong, Hong Kong
Fernando Pereira, Institute of Telecommunications, Portugal
Tony Pridmore, University of Nottingham, UK
Guoping Qiu, University of Nottingham, UK
Mark Sanderson, University of Sheffield, UK
Bertrand Le Saux, Ecole Normale Supérieure de Cachan, France
Gerald Schaefer, Aston University, UK
Raimondo Schettini, DISCO, University of Milano-Bicocca, Italy
Linlin Shen, Shenzhen University, China
Timothy Shih, Tamkang University, Taiwan
Yap-Peng Tan, Nanyang Technological University, Singapore
Qi Tian, University of Texas at San Antonio, USA
Martin Varley, University of Central Lancashire, UK
Giuliana Vitiello, University of Salerno, Italy
James Z. Wang, The Pennsylvania State University, USA
Roland Wilson, Warwick University, UK
Raymond Wong, National ICT Australia, Australia
Fei Wu, Zhejiang University, China
Jian Kang Wu, Institute for Infocomm Research, Singapore
Yihong Wu, Institute of Automation, Chinese Academy of Science, China
Xiang-Yang Xue, Fudan University, China
Pong Chi Yuen, Hong Kong Baptist University, Hong Kong
Matthias Zeppelzauer, Vienna University of Technology, Austria
Zhi-Hua Zhou, Nanjing University, China

Table of Contents

Visual Data Mining

Ubiquitous and Mobile Visual Information Systems

Semantics

2D/3D Graphical Visual Data Retrieval

Applications of Visual Information Systems

Visual Information Retrieval – Future Directions and Grand Challenges

Michael Lew

LIACS Media Lab, Leiden University
mlew@liacs.nl

Abstract. We are at the beginning of the digital Age of Information, a digital Renaissance allowing us to communicate, share, and learn in novel ways and resulting in the creation of new paradigms. However, having access to all of the knowledge in the world is pointless without a means to search for it. Visual information retrieval is poised to give access to the myriad forms of images and video, comprising knowledge from individuals and cultures to scientific fields and artistic communities. In this paper I summarize the current state of the field and discuss promising future directions and grand challenges.

Keywords: Visual Information Retrieval, Grand Challenges.

1 Visual Information Retrieval

Millennia ago, the Egyptians created the Library of Alexandria, an attempt to collect all of the knowledge in the world and store it in one vast library. The people who were responsible for collecting, indexing, and storing the knowledge were the earliest members of our field. They had the challenge of preserving the knowledge of their culture for future generations. This paper is a summary of the recent work in the field of Visual Information Retrieval (VIR) as described in the survey found in [1].

Today, we live in a world flooded with limitless data from every corner of the globe. The goal of the field of Visual Information Retrieval (VIR) is to develop new paradigms and theories for how to collect, store, analyze, search, and summarize visual information [2,3]. Raw bits are not enough. We must convert the bits into semantic concepts, meaningful translations of the data, and thereby bridge the semantic gap between the computer and humans.

Even though VIR encompasses many areas such as new features, selection algorithms, and similarity measures [11,12,13] and aspects such as high performance data structures and algorithms for very large data repositories, the focus in this article is on human centered aspects which are crucial to the future success of VIR. Next we discuss visual concept detection.

Early systems for face detection came from the field of Computer Vision and were limited to common assumptions such as (1) simple background such as a single color; (2) only one face per image; (3) no facial expressions; (4) no glasses; and (5) no occlusions over any part of the face. While the early face detection systems [16] had intriguing theory underlying the methods, the systems were not useful in VIR because

G. Qiu et al. (Eds.): VISUAL 2007, LNCS 4781, pp. 1–4, 2007.

the assumptions were too strict. In the past ten years major leaps forward have occurred as our field has addressed the difficult challenge of detecting objects in complex backgrounds. By the mid 90s, VIR researchers had succeeded in creating robust face detection systems founded on new paradigms such as information theory [7,8] which had eliminated the previous limitations and assumptions.

The next step was to generalize the face detector system to different kinds of visual concepts. Instead of only detecting faces, one would want to detect trees, sky, etc. Using similar methods as in the detection of faces, by the early 21st century, researchers also created robust systems for detection of simple visual concepts [4-6] such as sky, water, trees, and rocks.

The importance of the detection of visual concepts can not be understated. By automatically detecting visual concepts in imagery, we are directly bridging the semantic gap, bringing meaning to raw or senseless data. The visual concepts are the words upon which we can then build languages to describe and query for knowledge.

While research was indeed progressing forward, our field was still in the early stages of formation. In the late 90s conferences such as the *International Conference on Visual Information Systems* (VISUAL) were primarily focused on scientific researchers sharing their work with other researchers. To make our field stronger it was felt that we had to address at least two new challenges. First, it was clear that we needed researchers and practitioners to share their collective knowledge. Second, although we had succeeded in creating several promising systems, it was not clear how to perform quantitative evaluation. How could we scientifically say that one system outperformed another system or compare systems? These two challenges sparked the founding of several important meetings, notably, the International Conference on Image and Video Retrieval (CIVR) [3,14] and NIST TRECVID [15]. The mission of CIVR was *"to illuminate the state of the art in image and video retrieval between researchers and practitioners throughout the world."* The goal of TRECVID was *"to encourage research in information retrieval by providing a large test collection, uniform scoring procedures, and a forum for organizations interested in comparing their results."*

2 Current Research

Thus, in the early 21st century we were at a stage where Visual Information Retrieval had made substantial progress [1], but there were still major problems ahead. Beyond detecting faces, it is important to consider the temporal aspect such as understanding facial expressions such as emotional states like *happy* or *sad* in temporal sequences [9,10], or more generally, detecting visual concepts in time dependent imagery. The early visual concept detectors also had accuracy deterioration as the number of concepts grew. Reasonable results could be found at ten to twenty concepts, but not higher. If we could develop a visual concept detector which could grasp a thousand concepts, that might be sufficient for the development of a general language for describing the world.

In the area of interactive search [1], considerable research has addressed the algorithmic and learning issues, which have included varying feedback ratings, different architectures for well known learning algorithms such as Support Vector

Machines and Neural Networks, integration of multiple modalities, etc. However, few have addressed the problem of finding subimages which can occur when a user is interested in one part of an image, but not the rest. Furthermore, human users do not want to manually classify hundreds to thousands of examples for every search session, which has been called the small training set problem - How can we optimize the interactive search process for a small amount of interaction feedback?

Overall, recent research has been quite promising in terms of both incremental advances and in proposing new theories and paradigms. One of the next anticipated steps is for contributions in content based visual information retrieval to be used in a widely used application such as the next Google. Toward that end, we as a community would need to continue our work toward robust systems which work effectively under a wide range of real world imagery.

3 Grand Challenges

Current visual concept detection algorithms are only effective on small numbers of visual concepts, which can be said to be promising but is certainly insufficient for a generally useful vocabulary. Therefore, the first grand challenge is to create visual concept detectors which work robustly on hundreds of visual concepts instead of twenty.

The second grand challenge is to perform multi-modal analysis which exploits the synergy between the various media and uses knowledge sources in a deeper way. An example would be Wikipedia as an extensive knowledge source which we could tap into for fundamental knowledge of the world.

Beyond browsing and searching, how do we develop systems to allow users to gain insight and education? This is the third challenge which is to develop systems which focus on human-centered interaction not just toward searching but also toward gaining insight and knowledge.

The fourth grand challenge stems from the core interactive dialog between a librarian and a user. In some way the librarian asks context dependent, relevant, and intelligent questions to determine what the user really wants. How can we achieve this deeper level of relevance feedback between computers and humans?

In the early 21st century, TRECVID was an excellent example of combining researchers and users toward scientific benchmarking and evaluation. It is frequently forgotten that test sets can be used not just for benchmarking, but also for improving a system, revealing insights into strengths and weaknesses. The fifth grand challenge is to develop test databases and situations with emphasis on truly representative test sets, usage patterns, and aiding the researcher in improving his algorithm. For example, how can we make a good test set for relevance feedback?

4 Final Remarks

Regarding the future, the most important strength of our community is the ongoing sharing of knowledge between researchers and practitioners. As long as researchers keep the systems centered on humans, VIR will continue to bring significant advances to the world.