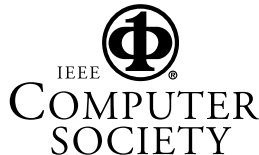


# ICME 2006

## Program Guide

### INTERNATIONAL CONFERENCE ON MULTIMEDIA AND EXPO

JULY 9 - 12, 2006  
HILTON TORONTO  
TORONTO, ONTARIO, CANADA



SPONSORED BY

THE INSTITUTE OF ELECTRICAL AND ELECTRONICS ENGINEERS  
CIRCUITS AND SYSTEMS SOCIETY  
COMMUNICATIONS SOCIETY  
COMPUTER SOCIETY  
SIGNAL PROCESSING SOCIETY

WITH SUPPORT FROM

Microsoft®  
**Research**



RYERSON UNIVERSITY

RYERSON UNIVERSITY

Engineering, Architecture & Science

## TABLE OF CONTENTS

---

CONFERENCE AT A GLANCE.....	3
PARALLEL TECHNICAL SESSIONS.....	4
MAPS.....	5
WELCOME TO ICME 2006 AND TORONTO.....	6
TECHNICAL PROGRAM OVERVIEW.....	7
ORGANIZING COMMITTEE.....	9
TECHNICAL PROGRAM COMMITTEE.....	10
SPECIAL SESSIONS.....	11
TPC MEMBERS.....	11
KEYNOTE TALKS.....	15
TUTORIALS.....	18
REGISTRATION.....	19
EXHIBITION.....	19
MEETING SCHEDULE.....	19
WELCOME TO TORONTO.....	20
CLIMATE IN TORONTO.....	20
ENTRY REQUIREMENTS.....	20
CUSTOMS INFORMATION.....	21
TORONTO AIRPORTS.....	21
RENTAL CAR INFORMATION.....	22
CURRENCY.....	22
TAXATION.....	23
GETTING AROUND.....	23
CONFERENCE VENUE & HOW TO GET THERE.....	24
NEIGHBORHOODS.....	24
PLACES TO EXPLORE.....	26
DAY TRIPS.....	27
TIPPING.....	28
HELPFUL PHONE NUMBERS.....	28
TECHNICAL PROGRAM.....	29
AUTHOR INDEX.....	175

# CONFERENCE AT A GLANCE

---

All events to be held in the Hilton Toronto except the banquet, which will be held in the CN Tower.

## SUNDAY, JULY 9, 2006

- 08:30 - 17:00 Registration, Foyer, Concourse Level
- 09:00 - 12:00 Tutorials, Concourse Level
- 10:30 - 11:00 Coffee Break
- 13:30 - 16:30 Tutorials, Concourse Level
- 15:00 - 15:30 Coffee Break
- 18:00 - 20:00 Welcome Reception, Toronto Ballroom, Concourse Level

## MONDAY, JULY 10, 2006

- 07:00 - 17:30 Registration, Foyer, Concourse Level
- 07:50 - 08:00 Opening Ceremony, Toronto I/II Ballroom, Concourse Level
- 08:00 - 09:00 Keynote Presentation, Toronto I/II Ballroom, Concourse Level
- 09:00 - 09:30 Coffee Break, Foyer, Concourse Level
- 09:30 - 10:50 Parallel Oral Presentation Sessions (MA1-L), Concourse Level
- 09:30 - 10:30 Parallel Poster Presentation Sessions (MA1-P), Concourse Level
- 10:50 - 12:10 Parallel Oral Presentation Sessions (MA2-L), Concourse Level
- 10:50 - 11:50 Parallel Poster Presentation Sessions (MA2-P), Concourse Level
- 12:10 - 13:10 Lunch Break (on your own)
- 13:10 - 14:30 Parallel Oral Presentation Sessions (MP1-L), Concourse Level
- 13:30 - 14:30 Parallel Poster Presentation Sessions (MP1-P), Concourse Level
- 14:30 - 15:50 Parallel Oral Presentation Sessions (MP2-L), Concourse Level
- 15:50 - 16:20 Coffee Break, Foyer, Concourse Level
- 16:20 - 17:20 Keynote Presentation, Toronto I/II Ballroom, Concourse Level

## TUESDAY, JULY 11, 2006

- 08:00 - 16:00 Registration, Foyer, Concourse Level
- 08:00 - 09:00 Keynote Presentation, Toronto I/II Ballroom, Concourse Level
- 09:00 - 09:30 Coffee Break, Foyer, Concourse Level
- 09:30 - 10:50 Parallel Oral Presentation Sessions (TA1-L), Concourse Level
- 09:30 - 10:30 Parallel Poster Presentation Sessions (TA1-P), Concourse Level
- 10:50 - 12:10 Parallel Oral Presentation Sessions (TA2-L), Concourse Level
- 10:50 - 11:50 Parallel Poster Presentation Sessions (TA2-P), Concourse Level
- 12:10 - 13:10 Lunch Break (on your own)
- 13:10 - 14:30 Parallel Oral Presentation Sessions (TP1-L), Concourse Level
- 13:30 - 14:30 Parallel Poster Presentation Sessions (TP1-P), Concourse Level
- 14:30 - 15:50 Parallel Oral Presentation Sessions (TP2-L), Concourse Level
- 14:50 - 15:50 Parallel Poster Presentation Sessions (TP2-P), Concourse Level
- 16:00 - 17:45 Cocktail Reception, Horizons Café, **CN Tower**
- 17:45 Conference Banquet, 360 Restaurant, **CN Tower**

## WEDNESDAY, JULY 12, 2006

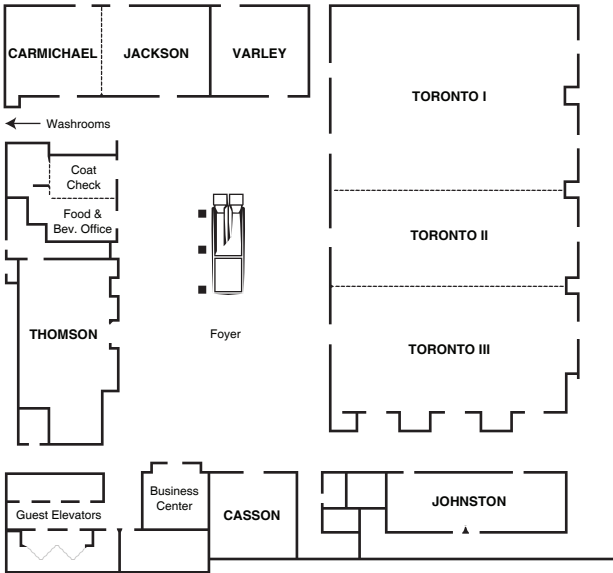
- 08:00 - 16:00 Registration, Foyer, Concourse Level
- 08:00 - 09:00 Keynote Presentation, Toronto I/II Ballroom, Concourse Level
- 09:00 - 09:20 Awards Presentation
- 09:20 - 09:50 Coffee Break, Foyer, Concourse Level
- 09:50 - 11:10 Parallel Oral Presentation Sessions (WA1-L), Concourse Level
- 11:10 - 12:30 Parallel Oral Presentation Sessions (WA2-L), Concourse Level
- 12:30 - 13:40 Lunch Break (on your own)
- 13:40 - 15:00 Parallel Oral Presentation Sessions (WP1-L), Concourse Level
- 15:00 - 15:30 Coffee Break, Foyer, Concourse Level
- 15:30 - 16:50 Parallel Oral Presentation Sessions (WP2-L), Concourse Level

# PARALLEL TECHNICAL SESSIONS

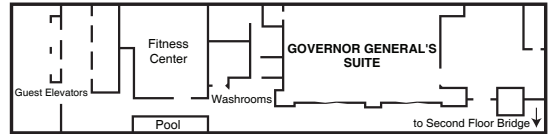
Time	Carmichael	Johnston	Thomson	Toronto I	Toronto II	Varley
Mon, July 10 09:30 - 10:50	MA1-L1 Feature Extraction and Segmentation I	MA1-L2 Processors & Multimedia I	MA1-L3 Content Understanding I	MA1-L4 Transcoding X to H.264 I	MA1-L5 Multimedia Interfaces I	MA1-L6 Multimedia Database Applications I
Mon, July 10 09:30 - 10:30	<b>Poster Session Toronto III</b>	MA1-P1 Media Coding	MA1-P2 Media Coding and Processing II	MA1-P3 Stereo, Multiview and Media Analysis	MA1-P4 Media Coding and Processing I	
Mon, July 10 10:50 - 12:10	MA2-L1 Feature Extraction and Segmentation II	MA2-L2 Processors & Multimedia II	MA2-L3 Content Understanding II	MA2-L4 Transcoding X to H.264 II	MA2-L5 Multimedia Interfaces II	MA2-L6 Multimedia Database Applications II
Mon, July 10 10:50 - 11:50	<b>Poster Session Toronto III</b>	MA2-P1 VLSI System-on-Chip Integrated Multimedia Systems	MA2-P2 Multimedia Interfaces: Searching and Browsing	MA2-P3 Multimedia Interfaces III	MA2-P4 Emerging Technologies for Next-Generation Multimedia Systems: Challenges and Opportunities	MA2-P5 Media Signal Processing I
Mon, July 10 13:10 - 14:30	MP1-L1 Feature Extraction and Classification I	MP1-L2 Perceptual Processing I	MP1-L3 Content Understanding III	MP1-L4 Authentication and Forensic Analysis	MP1-L5 P2P Multimedia I	MP1-L6 Multimedia Computing and Software I
Mon, July 10 13:30 - 14:30	<b>Poster Session Toronto III</b>	MP1-P1 Wireless and Mobile Multimedia	MP1-P2 Multimedia Applications I	MP1-P3 H.264 II	MP1-P4 Coding and Compression	
Mon, July 10 14:30 - 15:50	MP2-L1 Object Tracking, Indexing and Searching	MP2-L2 Perceptual Processing II	MP2-L3 Content Understanding IV	MP2-L4 Coding Algorithms and Standards I	MP2-L5 P2P Multimedia II	MP2-L6 Media Signal Processing II
Tue, July 11 09:30 - 10:50	TA1-L1 Digital Watermarking, Data Hiding and Steganography I	TA1-L2 Multi-User Collaboration and Cross Layer Optimization in Wireless Multimedia Communication I	TA1-L3 Event and Activity Recognition I	TA1-L4 Audio and Speech Processing I	TA1-L5 Peer-to-Peer Multimedia Streaming	TA1-L6 Interactive Multimedia Content Analysis
Tue, July 11 09:30 - 10:30	<b>Poster Session Toronto III</b>	TA1-P1 Content Understanding V	TA1-P2 Object Recognition and Tracking	TA1-P3 Other Applications	TA1-P4 Multimedia Computing and Software II	TA1-P5 Analysis and Synthesis of Speech and Audio I
Tue, July 11 10:50 - 12:10	TA2-L1 Digital Watermarking, Data Hiding and Steganography II	TA2-L2 Multi-User Collaboration and Cross Layer Optimization in Wireless Multimedia Communication II	TA2-L3 Multimedia in WWW	TA2-L4 Audio and Speech Processing II	TA2-L5 Multimedia Adaptation and Cross-Layer QoS for Wireless Networks	TA2-L6 Media Signal Processing III
Tue, July 11 10:50 - 11:50	<b>Poster Session Toronto III</b>	TA2-P1 Multimedia Networking	TA2-P2 Multimedia Systems, Architectures and Hardware	TA2-P3 Multimedia Applications II	TA2-P4 Authentication, Data Hiding and Others	
Tue, July 11 13:10 - 14:30	TP1-L1 Digital Watermarking, Data Hiding and Steganography III	TP1-L2 Analysis and Synthesis of Speech and Audio II	TP1-L3 Advances in Networked Video	TP1-L4 P2P Multimedia III	TP1-L5 Joint Source and Channel Coding for Wireless Multimedia	TP1-L6 Advanced Media Coding I
Tue, July 11 13:30 - 14:30	<b>Poster Session Toronto III</b>	TP1-P1 Event and Activity Recognition II	TP1-P2 Multimedia Networking	TP1-P3 Face Detection	TP1-P4 Digital Watermarking	TP1-P5 Applications to Sports Video
Tue, July 11 14:30 - 15:50	TP2-L1 Object Recognition	TP2-L2 Multimedia Interfaces: Speech and Audio	TP2-L3 Coding Algorithms and Standards II	TP2-L4 P2P Multimedia IV	TP2-L5 3D-TV: Primed for Success? I	TP2-L6 H.264 I
Tue, July 11 14:50 - 15:50	<b>Poster Session Toronto III</b>	TP2-P1 Feature Extraction and Classification II	TP2-P2 Multimedia Database Query Processing and Retrieval	TP2-P3 Segmentation		
Wed, July 12 09:50 - 11:10	WA1-L1 Quality of Service Support for Multimedia Transport	WA1-L2 Media Processing and Algorithms	WA1-L3 3D-TV: Primed for Success? II	WA1-L4 Applications to Sports Video	WA1-L5 Feature Extraction and Segmentation III	WA1-L6 Coding Algorithms and Standards III
Wed, July 12 11:10 - 12:30	WA2-L1 Application Driven Multimedia Systems and Architectures	WA2-L2 Multimedia Systems Engineering: System Modeling, Algorithm, Simulation, and Analysis	WA2-L3 Multimedia over Wireless Networks	WA2-L4 Music Indexing and Retrieval	WA2-L5 Multimedia for Healthcare	WA2-L6 Coding Algorithms and Standards IV
Wed, July 12 13:40 - 15:00	WP1-L1 Mobile Multimedia Applications I	WP1-L2 Programmable Hardware Platform for Multimedia Systems and Applications	WP1-L3 Stereo, Multiview and 3D	WP1-L4 Multimedia Technologies for Distance Learning	WP1-L5 Feature Extraction and Segmentation IV	WP1-L6 Multimedia for Virtual Reality
Wed, July 12 15:30 - 16:50	WP2-L1 Mobile Multimedia Applications II	WP2-L2 Protocols for Multimedia Delivery Over Networks	WP2-L3 3D-TV: Primed for Success? III	WP2-L4 Encryption and Access Control		

# MAPS

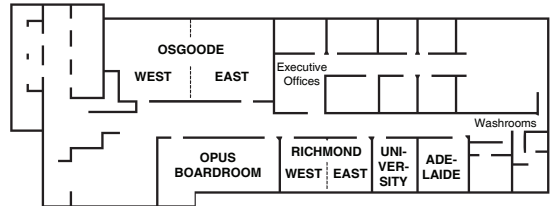
## HILTON TORONTO - CONVENTION LEVEL



## HILTON TORONTO - SECOND FLOOR



## HILTON TORONTO - THIRD FLOOR



## DOWNTOWN TORONTO



# WELCOME TO ICME 2006 AND TORONTO

---

On behalf of the Organizing Committee, it is our humble pleasure to invite/welcome you to Toronto and to the 2006 IEEE International Conference on Multimedia and Expo (ICME 2006). We consider ourselves fortunate to have had the opportunity to organize this 7th ICME and offer you a taste of Canadian hospitality.

We are thrilled to see that this year ICME draw a record number of 1022 paper submissions. We are fortunate to have the invaluable cooperation of our colleagues as Technical Program Committee (TPC) Co-Chairs, Professors Chang Wen Chen and Kostas Plataniotis, and TPC Vice Chairs representing the four participating IEEE Societies, Dr. Mike Christel (Computer), Dr. Jean-Luc Dugelay (Signal Processing), Professor Panos Nasiopoulos (Circuits and Systems) and Professor En-hui Yang (Communications), who put together a rich in content and high-quality technical program spanning three full days of six lecture and four poster sessions running in parallel and covering 10 different tracks, representing the technical activities in multimedia in the four Societies. In the program are included 10 special sessions and a three day exhibition.

On the day before the conference starts, we offer seven high quality tutorials on several important aspects of multimedia. The conference features four keynote speeches offered by our distinguished speakers: K. J. Ray Liu, on “Multimedia Forensics: Where Sherlock Holmes Meets Signal Processing”, Tas Venetsanopoulos, on “Multimedia Signal Processing and Content-Based Image Retrieval (a Survey)” Sun-Yuan Kung, on “Machine Learning for Multi-Modality Genomic Data Mining” and Thomas Huang, on “Ten Challenging Problems in Content-based Image and Video Retrieval”.

In addition to the technical program, we have also put together a social program that will at least keep you busy: a welcome reception in the hosting hotel, and the formal conference reception and banquet in 360 Restaurant on CN Tower, the tallest free standing structure on this planet.

We want to share with you our gratitude towards all members of the organizing committee for their dedication to the success of ICME 2006. Working like one person for a common goal in the demanding task of ICME organization made us proud that we got involved in this effort. For this, we thank them all.

We would like to thank Microsoft Research and IBM Research for their sponsorships to the Best Paper Award and Best Student Paper Award. We would also like to thank Ryerson Faculty of Engineering, Architecture and Science and Ryerson University for their generous support and army of volunteers.

We are very grateful to Conference Management Services, Inc, for making our work as organizers as easy as it could be and for offering their unique ideas and initiatives for us all to enjoy.

We would like to invite you to visit Toronto, your first time in Canada with ICME, especially at this time of the year when nature is welcoming everybody, and enjoy both the rich technical program and the endless opportunities to explore the place and the culture. We hope your stay here will be both rewarding and memorable.

Ling Guan, Hong-Jiang Zhang  
*ICME 2006 General Co-Chairs*

# TECHNICAL PROGRAM OVERVIEW

---

It is our great honor and pleasure to welcome you to the 2006 IEEE International Conference on Multimedia & Expo (ICME 2006) in Toronto, Canada. This year, the technical program has built on a great ICME tradition as the world's premier technical conference in the field of multimedia and has achieved record setting performance in paper submission, acceptance rate, and accepted papers.

We are pleased to report that we have received record number of submissions with 1022 papers from 54 different countries. This record number of submissions shows that ICME has already become the conference of choice for all fields of multimedia. With such a significant increase in paper submissions, we have to recruit over 300 experts to form the technical program committee (TPC) to ensure that all papers are reviewed by at least three members of the TPC. The members of the TPC listed in the following pages include experts from all four sponsoring society: Circuits and Systems, Communications, Computer, and Signal Processing. These experts generously volunteered their time and effort to ensure the excellence of the technical program and to handle the significant increase of paper submissions. On behalf of the ICME 2006 Organizing Committee, we would like to thank all TPC members for their dedication and timely review of the record number of submissions.

Based on the reviews provided by the Technical Program Committee, the Technical Program Chairs selected 552 papers that were organized into 51 oral sessions, 30 poster sessions, and 10 special sessions. The accepted papers include special session papers that have much higher acceptance rate. Even though the number of the accepted papers sets a record, the acceptance rate from general submission stands at record low of 51%. Therefore, this year's ICME has been the most selective ICME of all. Among the almost 49% rejected papers from general submissions, many valuable technical works had to be declined due to time and space constraints. On behalf of the Organizing Committee, we express our regret to those authors whose submissions could not be accommodated at ICME 2006. We extend warm welcome to these authors to join us at ICME 2006 to exchange with your colleagues at the conference. Overall, the process resulted in an excellent technical program covering a wide range of topics, including several emerging fields in multimedia.

The highlights of this year's ICME are the keynotes speeches given by four world renowned experts in multimedia. On Monday July 10, we will have Prof. K. J. Ray Liu from the University of Maryland delivering a keynote speech on "Multimedia Forensics: Where Sherlock Holmes Meets Hollywood" in the morning, and Prof. Tas Venetsanopoulos from the University of Toronto delivering a keynote speech on "Multimedia Signal Processing and Content-Based Image Retrieval (a Survey)" in the afternoon. On Tuesday July 11, we will have Prof. Sun-Yuan Kung from Princeton University delivering a keynote speech on "Machine Learning for Multi-Modality Genomic Data Mining" and on Wednesday July 12, we will have Prof. Thomas Huang from the University of Illinois at Urbana-Champaign delivering a keynote speech on "Ten Challenging Problems in Content-based Image and Video Retrieval." We encourage every participant of ICME 2006 to attend all four outstanding keynote speeches.

The technical activity of ICME 2006 will officially start on Sunday, July 9, with seven tutorials organized by the Tutorials Co-Chairs, Liang-Gee Chen and Anthony Vetro. Attendees can get overviews of the state of the art in several key areas of multimedia:

1. Steganography and Steganalysis: Fundamentals, Algorithms and Future Research, by Prof. Yun Q. Shi, New Jersey Institute of Technology
2. Peer-to-Peer Multimedia Applications, by Dr. Jin Li,
3. Digital Inpainting, by Prof. Timothy K. Shih
4. Capturing Light and Motion: Multimedia Techniques for Animation and Movie Production, by Prof. Howard Leung and Prof. Tsuhan Chen
5. Hardware and Software Techniques for Multimedia Systems, by Prof. Nikil Dutt and Prof. Wayne Wolf

## 6. Music Information Retrieval, by Prof. George Tzanetakis

Ten special sessions have been organized by the Special Sessions Chairs, Jiebo Luo and Andre Zaccarin. Special sessions were selected based on the quality of the proposals, timeliness of the subject and qualification of the authors. The special sessions are:

1. Processor & Multimedia, by Yen-Kuang Chen and Eric Debes
2. Transcoding to H.264 by Hari Kalva
3. Advances in Networked Video, by Mary Comer and Xiaojun Lin
4. Perceptual Visual Processing, by Xiaokang Yang, Weisi Lin, and Minoru Etoh
5. 3-D TV: Primed for Success? by Liang Zhang, Wa James Tam, and Hirokazu Yamanoue
6. Multi-User Collaboration and Cross Layer Optimization in Wireless Multimedia Communication, by Zhu Li, Pascal Frossard, and Aggelos Katsaggelos
7. Interactive Multimedia Content Analysis and Applications, by Xian-Sheng Hua and Qi Tian, University of Texas, San Antonio
8. P2P Multimedia, by Deepa Kundur, Zhu Liu, Madjiid Merabti, and Heather Yu
9. Multimedia Technologies for Distance Learning, by Qing Li and Timothy K. Shih
10. Multimedia Technologies for Healthcare, by Datong Chen, Howard D. Watclar, and Jie Yang

We have worked very hard to organize the sessions coherently in order to minimize the potential overlap. We are confident that the attendees will find a great deal of interesting and stimulating material in the technical program to keep them challenged and engaged over the three days of the technical program and beyond. Undoubtedly, organizing a conference with record setting submission as ICME 2006 is a challenging and demanding task. We will not be able to successfully accomplish such task without numerous volunteers' help. We wish to thank the members of the Organizing Committee for the tremendous amount of time and effort spent into this process, making joint decisions and monitoring the process along the way. In particular, we would like to express our deep appreciation to the TPC Vice Co-Chairs, Mike Christel of Carnegie Mellon University, Jean-Luc Dugelay of Eurecom, France, Panos Nasiopoulos of University of British Columbia, and En-hui Yang of University of Waterloo, for helping us in the entire process of review, acceptance decision, and session organization. We are also greatly in debt to all the Track Chairs for their dedicated work in the tremendous review process. It is these Track Chairs who actually recruited outstanding team of expert reviewers to carry out the challenging tasks of reviewing record number of papers for ICME 2006. Their contributions to ICME 2006 are indeed invaluable.

We would like to thank Microsoft Research for sponsoring the ICME Best Paper Award and IBM for sponsoring the ICME Best Student Paper Award. Their generous support in sponsoring these awards will inspire more innovations in the field of multimedia and will encourage more students to join this young and vibrant multimedia community to pursue their professional career.

We wish to acknowledge the able and tireless contribution of Lance Cotton and his team at Conference Management Services, Inc for managing the online review process, assembling the technical program, and accommodating the selection process for the Best Paper Award and the Best Student Paper Award.

Finally, our appreciation goes to the thousands of multimedia researchers worldwide who have contributed their excellent work to the outstanding technical program of ICME 2006. We hope that you will enjoy the conference and find the technical program exciting. We also hope that you will find some time to attend the social program of ICME 2006 and to visit the great city of Toronto.

Chang Wen Chen and Kostas N. Plataniotis  
*ICME 2006 Technical Program Co-Chairs*



# ORGANIZING COMMITTEE

---

## General Co-Chairs

Ling Guan

*Ryerson University*

Hong-Jiang Zhang

*Microsoft Research Asia*

## Technical Program Co-Chairs

Chang Wen Chen

*Florida Institute of Technology*

Kostas. N. Plataniotis

*University of Toronto*

## Technical Program Vice Co-Chairs

Mike Christel

*Carnegie Mellon University*

Jean-Luc Dugelay

*Eurecom*

Panos Nasiopoulos

*University of British Columbia*

En-hui Yang

*University of Waterloo*

## Finance Chair

Liping Fang

*Ryerson University*

## Special Session Co-Chairs

Jiebo Luo

*Eastman Kodak*

André Zaccarin

*Université Laval*

## Tutorial Co-Chairs

Liang-Gee Chen

*National Taiwan University*

Anthony Vetro

*Mitsubishi Electric Research Lab*

## Local Arrangement Co-Chairs

Alagan Anpalagan

*Ryerson University*

Sri Krishnan

*Ryerson University*

## Publicity Co-Chairs

Robert Dony

*University of Guelph*

Xiao-Ping Zhang

*Ryerson University*

## Publication Co-Chairs

Amir Asif

*York University*

Liang Zhang

*Communications Research Centre Canada*

## Demo/Exhibit Chair

Dimitri Androutsos

*Ryerson University*

## Registration Co-Chairs

Ivan Lee

*Ryerson University*

Paisarn Muneesawang

*UAE University*

## Asian Liaison

Horace Ip

*City University of Hong Kong*

## European Liaison

Joern Ostermann

*University of Hanover*

## Webmaster

Jose Lay

*University of Sydney*

## Conference Managed by

Conference Management Services, Inc.

# TECHNICAL PROGRAM COMMITTEE

---

## Technical Program Co-Chairs

Chang Wen Chen  
*Florida Institute of Technology*  
Kostas. N. Plataniotis  
*University of Toronto*

## Technical Program Vice Co-Chairs

Mike Christel  
*Carnegie Mellon University*  
Jean-Luc Dugelay  
*Eurecom*  
Panos Nasiopoulos  
*University of British Columbia*  
En-hui Yang  
*University of Waterloo*

## Multimedia Content Analysis

Dr. Ching-Yung Lin  
*IBM*  
Prof. Qi Tian  
*University of Texas, San Antonio*  
Dr. Adriana Dumitras  
*Apple Computer*

## Multimedia Signal Coding and Processing

Dr. Shipeng Li  
*Microsoft Research*  
Prof. Touradj Ebrahimi  
*EPFL*  
Prof. George Tzanetakis  
*University of Victoria*

## Multimedia Database

Prof. Guojun Lu  
*Monash University*  
Prof. B. S. Manjunath  
*University of California, Santa Barbara*

## Wireless and Mobile Multimedia

Prof. Reiner Creutzburg  
*Fachhochschule Brandenburg*  
Prof. Jianfei Cai  
*Nanyang Technological University*

## Multimedia Networking

Dr. Haitao Zheng  
*University of California, Santa Barbara*  
Prof. Dapeng Oliver Wu  
*University of Florida*

## Multimedia Security

Dr. Qibin Sun  
*I2R*  
Dr. Ronggong Song  
*National Research Council Canada*

## Multimedia Interface and Interaction

Dr. Andreas Girgensohn  
*FX Palo Alto Laboratory*  
Prof. Alan F. Smeaton  
*Dublin City University*

## Multimedia Hardware, Architecture and Systems

Dr. Wai-Chi Fang  
*NASA JPL, California Institute of Technology*

## Multimedia Computing and Software

Dr. David Doermann  
*University of Maryland*  
Prof. Aidong Zhang  
*SUNY at Buffalo*

## Multimedia Applications

Prof. Noboru Babaguchi  
*Osaka University*  
Prof. Borko Furht  
*Florida Atlantic University*

## SPECIAL SESSIONS

---

### **Processor & Multimedia**

Yen-Kuang Chen, Intel Corporation and Eric Debes, Intel Corporation

### **Transcoding to H.264**

Hari Kalva, Florida Atlantic University

### **Advances in Networked Video**

Mary Comer, Purdue University and Xiaojun Lin, Purdue University

### **Perceptual Visual Processing**

Xiaokang Yang, Shanghai Jiao Tong University, Weisi Lin, Institute for Infocomm Research and Minoru Etoh, Multimedia Laboratories, NTT DoCoMo

### **3-D TV: Primed for Success?**

Liang Zhang, Communications Research Centre Canada, Wa James Tam, Communications Research Centre Canada and Hirokazu Yamanoue, NHK Science & Technical Research Laboratories

### **Multi-User Collaboration and Cross Layer Optimization in Wireless Multimedia Communication**

Zhu Li, Motorola Labs, Pascal Frossard, EPFL and Aggelos Katsaggelos, Northwestern University

### **Interactive Multimedia Content Analysis and Applications**

Xian-Sheng Hua, Microsoft Research Asia and Qi Tian, University of Texas, San Antonio

### **P2P Multimedia**

Deepa Kundur, Texas A&M University, Zhu Liu, AT&T Research Lab, Madjid Merabti, Liverpool John Moores University and Heather Yu, Panasonic Information and Networking Technologies Laboratory

### **Multimedia Technologies for Distance Learning**

Qing Li, City University of Hong Kong and Timothy K. Shih, Tamkang University, Taiwan

### **Multimedia Technologies for Healthcare**

Datong Chen, Carnegie Mellon University, Howard D. Watclar, Carnegie Mellon University and Jie Yang, Carnegie Mellon University

## TPC MEMBERS

---

Rafeef Abugharbieh, University of British Columbia (ECE), Canada  
Tomasz Adamek, Dublin City University, Ireland  
Michael Adams, Dept. of ECE, University of Victoria  
John Adcock, FX Palo Alto Laboratory Inc  
Kiyoharu Aizawa, University of Tokyo, Japan  
Marios Angelides, Brunel University  
Rashid Ansari, University of Illinois/Chicago, USA  
John Apostolopoulos, Hewlett-Packard (HP) Labs, USA  
Yasuo Ariki, Kobe University, Japan  
David Atienza, IMEC & EPFL  
Oscar Au, Hong Kong University of Science and Technology, Hong Kong  
Ramazan Aygun, University of Alabama at Huntsville, USA  
Noboru Babaguchi, Osaka University  
Brian Bailey, University of Illinois  
Andrea Basso, University of Victoria, Canada  
Magdy Bayoumi, Univ of Louisiana at Lafayette, USA  
Mladen Berekovic, IMEC  
Susanne Boll, University of Oldenburg, Germany  
John Boreczky, FX Palo Alto Laboratory, USA  
Jill Boyce, Thomson  
Hua Cai, Microsoft Research Asia  
Jianfei Cai, Nanyang Technological University  
Kasim Selcuk Candan, Arizona State University, USA  
Jiannong Cao, Hong Kong Polytechnic University  
Lei Cao, The University of Mississippi  
Joe Carthy, University College Dublin  
Yee Sin Chan, University of Miami  
Chih-Wei Chang, Industrial Technology Research Institute  
Ee-Chien Chang, National University of Singapore, Singapore  
Jacob Charkareski, EPFL  
Lap-Pui Chau, Nanyang Technological University, Singapore  
Upendra Chaudhari, IBM T. J. Watson Research Center  
Datong Chen, Carnegie Mellon University  
Francine Chen, FX Palo Alto Laboratory  
Homer Chen, National Taiwan University  
John R. (Rong-Jian) Chen, National United University  
Liang-Gee Chen, National Taiwan University, Taiwan  
Oscal T.C. Chen, EE Dept of National Chung Cheng University  
Peisong Chen, Qualcomm  
Shu-Ching Chen, Florida International University, USA  
Tien-Fu Chen, CCU  
Trista Chen, Intel  
Yen-Kuang Chen, Intel Corporation  
Yih-Farn Chen, AT&T Labs - Research  
Corey Cheng, Dolby Laboratories  
Liang Cheng, NVIDIA  
Gene Cheung, Hewlett-Packard Laboratories Japan  
Liang-Tien Chia, Nanyang Technological University  
Jen-Shiun Chiang, Tamkang University  
Tihao Chiang, National Chiao-Tung University, Taiwan  
Shao-Yi Chien, National Taiwan University, Taiwan  
Patrick Chiu, FX Palo Alto Laboratory  
Reha Civanlar, Koc University, Turkey  
Mary Comer, Purdue University

Matthew Cooper, FX Palo Alto Laboratory, USA  
 Paulo Correia, Instituto Superior Tecnico  
 Jun-Hong Cui, University of Connecticut  
 Rozenn Dahyot, Trinity College Dublin, Ireland  
 Min Dai, QUALCOMM Incorporated  
 Pankaj Das, Department of Electrical and Computer Engineering, University of California at San Diego  
 Juan Carlos De Martin, Politecnico di Torino, Italy  
 Peter H.N. de With, University of Technology Eindhoven / LogicaCMG  
 Eric Debes, Intel Corporation  
 Laurent Denoue, FX Palo Alto Laboratory  
 Guiguang Ding, Tsinghua University  
 Chabane Djeraba, LIFL  
 David Doermann, University of Maryland  
 Adriana Dumitras, Apple Computer  
 Shahram Ebadollahi, IBM T.J. Watson Research Center, Hawthorn, NY  
 Touradj Ebrahimi, EPFL  
 Tomio Echigo, Osaka University, Japan  
 Bruce Elder, Ryerson University  
 Il Kyu Eom, School of Electronics, Electrical, and Communication Engineering, Pusan National University  
 Berna Erol, Ricoh California Research Center, USA  
 Minoru Etoh, NTT Docomo, Japan  
 Jianping Fan, UNC-Charlotte, USA  
 Wai-Chi Fang, NASA JPL, California Institute of Technology  
 Dirk Farin, Technical University Eindhoven (TU/e)  
 Sid Fels, University of British Columbia (ECE), Canada  
 Paul Ferguson, Dublin City University, Ireland  
 Xavier Fernando, Ryerson University  
 Chuan Heng Foh, Nanyang Technological University  
 Colum Foley, Dublin City University, Ireland  
 Pascal Frossard, EPFL  
 Orazio Gambino, Universita' degli Studi di Palermo  
 Deyun Gao, Nanyang Technological University  
 Wen Gao, Peking University, and Institute of Computing Technology, Chinese Academy of Sciences  
 Daniel Gatica-Perez, IDIAP Research Institute and Swiss Federal Institute of Technology, Lausanne (EPFL)  
 Georgina Gaughan, Dublin City University, Ireland  
 Werner Geyer, IBM Research  
 Jerry Gibson, UC Santa Barbara  
 Kwong Huang Goh, Institute for Infocomm Research  
 Abhijeet Golwelkar, Harris Corporation  
 Yihong Gong, NEC Labs America, USA  
 Richard Green, University of Canterbury  
 Ling Guan, Ryerson University  
 Huiping Guo, Department of Computer Science, California State University, USA  
 Jiun-In Guo, National Chung-Cheng University  
 Yang Guo, Thomson Research  
 Cathal Gurrin, Dublin City University, Ireland  
 Liu Hang, Thomson Inc  
 Yuwen He, Panasonic Singapore Laboratories  
 Yo-Sung Ho, Gwangju Institute of Science and Technology  
 Shih-Ta Hsiang, Motorola Labs  
 Hsu-Feng Hsiao, National Chiao-Tung University, Taiwan  
 Chiou-Ting Hsu, National Tsing Hua University, Taiwan  
 Winston Hsu, Columbia University, USA  
 Yu Hen Hu, University of Wisconsin-Madison  
 Xian-Sheng Hua, Microsoft Research Asia  
 Hsiang-Cheh Huang, National Kaohsiung Marine University  
 Jianwei Huang, Princeton University  
 Qingming Huang, JDL Lab, Graduate School, Chinese Academy of Sciences  
 Christopher Hughes, Intel  
 Ichiro Ide, Nagoya University, Japan  
 Horace Ip, City Univ. of Hong Kong  
 Horace Ip, City University of Hong Kong  
 Cyril Iskander, Florida Atlantic University  
 Rachel Jiang, Ryerson University  
 Chun Jin, Carnegie Mellon University  
 Jesse Jin, University of Newcastle, Australia  
 Wu Jonathan, University of Windsor  
 Gareth Jones, Dublin City University, Ireland  
 Joemon Jose, Glasgow University  
 Dan Jurca, Ecole Polytechnique Federale de Lausanne  
 Hari Kalva, Florida Atlantic University  
 Yoshinari Kameda, University of Tsukuba, Japan  
 Mohan Kankanhalli, National University of Singapore  
 Lance Kaplan, U.S. Army Research Laboratory, USA  
 Gour Karmakar, Monash University, Australia  
 Andreas Kassler, Karlstad University  
 John Kender, Columbia University  
 Lyndon Kennedy, Columbia University  
 Hua Kien, University of Central Florida, USA  
 Andrew Kinane, Dublin City University, Ireland  
 Lisimachos Paul Kondi, SUNY at Buffalo  
 Jiejun Kong, University of California, Los Angeles  
 Markus Koskela, Dublin City University  
 Alex Kot, Nanyang Technological University, Singapore  
 C.C. Jay Kuo, Dept of Electrical Engineering Systems, Univ of Southern CA  
 Azadeh Kushki, University of Toronto  
 Chunrong Lai, Intel  
 Wei Lai, Microsoft Research Asia  
 Kenneth Lam, Hong Kong Polytechnic University, Hong Kong  
 Rosa C. Lancini, CEFRIEL-Politecnico di Milano  
 Li Lao, University of California, Los Angeles  
 Eddie Law, Ryerson University  
 Herve Le Borgne, Dublin City University, Ireland  
 Gwo Giun (Chris) Lee, National Cheng Kung University, Tainan, Taiwan  
 Heung-Kyu Lee, Dept of EECS, KAIST  
 Hyowon Lee, Dublin City University  
 Victor Lee, Intel  
 Bart Lehan, Dublin City University, Ireland  
 Clement Leung, Victoria University  
 Elvis Leung, City University of Hong Kong  
 Howard Leung, City University of Hong Kong  
 Chung-Sheng Li, IBM T.J. Watson Research Center, Yorktown Heights, NY  
 Danjue Li, University of California at Davis  
 Eric Q Li, Intel  
 Houqiang Li, University of Science & Technology of China  
 Jiang Li, Microsoft Research Asia  
 Jin Li, Microsoft Research, USA  
 Qi Li, SDU  
 Shipeng Li, Microsoft Research Asia (China)

Tao Li, Florida International University  
 Wanqing Li, University of Wollongong  
 Wenlong Li, Intel  
 Xin Li, West Virginia University  
 Xuelong Li, University of London, United Kingdom  
 Xun Li, Celestial (Beijing) Semiconductor, Inc.  
 Ying Li, IBM T.J. Watson Research Center, Yorktown Heights, NY  
 Zhengguo Li, I2r, Singapore  
 Zhu Li, Multimedia Research Lab (MRL), Motorola Labs  
 Jie Liang, Simon Fraser University  
 Mark Liao, Institute of Applied Science and Engineering Research (Taiwan)  
 Wen-Nung Lie, National Chung Cheng University  
 Joo Hwee Lim, Institute for Infocomm Research, Singapore  
 Chia-Wen Lin, National Chung Cheng University  
 Xiaojun Lin, Purdue University  
 Xinggang Lin, Tsinghua University, China  
 Bin-Da (Brian) Liu, National Cheng Kung University  
 Jiangchuan(JC) Liu, University of Simon Fraser  
 Lurng-Kuo Liu, IBM  
 Qiong Liu, Fuji-Xerox Palo Alto Lab (FXPAL), USA  
 Wenyin Liu, City Uni. of HK  
 Zhengye Liu, Polytechnic University  
 Zhu Liu, AT&T Research Lab  
 Jian-Guang Lou, Microsoft Research Asia  
 Chun-Shien Lu, Academia Sinica  
 Lie Lu, Microsoft  
 Xiaolan Lu, Thomson Inc.  
 Yan Lu, Microsoft Research Asia  
 Rastislav LUKAC, University of Toronto  
 Chong Luo, Microsoft Research Asia  
 Yijing Luo, Motorola  
 Kai-Kuang Ma, Nanyang Technological University, Singapore  
 Wei-Ying Ma, Microsoft Research Asia (China)  
 Enrico Magli, Politecnico di Torino, Italy  
 Oge Marques, Florida Atlantic University  
 Karl MARTIN, University of Toronto  
 Kieran McDonald, Dublin City University, Ireland  
 Sin?ad McGivney, Dublin City University, Ireland  
 Michael McGuire, University of Victoria  
 Michael McHugh, Dublin City University, Ireland  
 John McKenna, Dublin City University, Ireland  
 Helen Meng, Chinese University of Hong Kong  
 Manzur Murshed, Monash University, Australia  
 Mor Naaman, Yahoo! Research Berkeley  
 Frank Nack, V2\_, Institute for the Unstable Media  
 Masahide Naemura, Advanced Telecommunications Research Institute International  
 Kumiyo Nakakoji, University of Tokyo  
 Yuichi Nakamura, Kyoto University  
 Apostol Natsev, IBM Thomas J. Watson Research Center  
 Eamonn Newman, University College Dublin  
 Tian-Tsong Ng, Columbia University  
 Chong Wah NGO, City University of Hong Kong  
 Conor Nugent, Trinity College Dublin, Ireland  
 Noel O'Connor, Dublin City University, Ireland  
 Neil O'Hare, Dublin City University, Ireland  
 Jens-Rainer Ohm, RWTH Aachen University  
 Timo Ojala, University of Oulu, Finland  
 Ee Ping Ong, Institute For Infocomm Research  
 Wei Tsang Ooi, National University of Singapore, Singapore  
 Antonio Ortega, Dept of Electrical Engineering Systems, Univ of Southern CA  
 Peshala Pahalawatta, Northwestern University  
 Jeng-Shyang Pan, National Kaohsiung University of Applied Sciences  
 Navneet Panda, University of California Santa Barbara  
 Chunyi Peng, Microsoft Research Asia  
 Ya-Ti Peng, University of Washington  
 Fernando Pereira, Instituto Superior Tecnico, Portugal  
 Chatzimisios Periklis, TEI of Thessaloniki  
 Adi Pinhas, Intel  
 Peter Pirsch, University of Hannover, Institut fuer Mikroelektronische Systeme  
 Konstantinos N. Plataniotis, University of Toronto  
 Benjamin Premkumar, SCE, Nanyang Tech University  
 Gao Qigang, Dalhousie University  
 Guoping Qiu, University of Nottingham, United Kingdom  
 Wei Qu, University of Illinois at Chicago  
 Majid Rabbani, Eastman Kodak Research Labs, USA  
 Regunathan Radhakrishnan, Mitsubishi Electrical Research Lab, USA  
 Susanto Rahardja, Institute for Infocomm Research, Singapore  
 Deepu Rajan, Nanyang Technological University, Singapore  
 Mika Rautiainen, University of Oulu, Finland  
 YongMan Ro, ICU  
 Sandra Rothwell, Centre for Digital Video Processing, Dublin City University  
 Sebastien Roux, France Telecom  
 Sujoy Roy, Institute for Infocomm Research  
 Shanq-Jang Ruan, National Taiwan University of Science and Technology  
 Stefan Rueger, Imperial College London  
 David Sadlier, Dublin City University, Ireland  
 Amir Said, Hewlett-Packard Labs, USA  
 Hideo Saito, Keio University  
 Andrew Salway, University of Surrey  
 Shin'ichi Satoh, National Institute of Informatics, Japan  
 Sorin Sav, Dublin City University, Ireland  
 Dan Schonfeld, University of Illinois at Chicago  
 Nicu Sebe, University of Amsterdam, The Netherlands  
 Shirani Shahram, McMaster University, Canada  
 Huai-Rong Shao, Samsung Electronics  
 Jacky Shen, Microsoft Research Asia  
 Sheng Mei Shen, Panasonic Singapore Laboratories  
 Sherman Shen, Dept of Electrical and Computer Engineering, Univ. of Waterloo, Canada  
 Yanming Shen, Polytechnic University  
 Yun-Qing Shi, New Jersey Institute of Technology, USA  
 Timothy K. Shih, Tamkang University, Taipei  
 Frank Shipman, Texas A&M University  
 Mei-Ling Shyu, University of Miami, USA  
 Cees Snoek, University of Amsterdam, The Netherlands  
 Hwangjun Song, POSTECH (Pohang Univ. of Science and Technology), Korea  
 Ronggong Song, National Research Council, Canada  
 Xiaodan Song, University of Washington  
 Yan Song, USTC  
 Yuqing Song, University of Michigan at Dearborn, USA

Eckehard Steinbach, Munich University of Technology  
 Scott Stevens, Carnegie Mellon University, USA  
 Thomas Stockhammer, Nomor Research, Munich, Germany  
 Lifeng Sun, Tsinghua University  
 Ming-Ting Sun, University of Washington, USA  
 Xiaoyan Sun, Microsoft Research Asia  
 Zhaohui Sun, GE Global Research  
 Wa James Tam, Communications Research Centre Canada  
 Wai-tian Tan, Hewlett Packard Company  
 Cheng-Yuan Tang, Huafan University  
 Dayong Tao, Nanyang Technological University  
 Wallapak Tavanapong, Iowa State University, USA  
 Clark N. Taylor, Brigham Young University  
 A. Murat Tekalp, Koc University  
 Shyh Wei Teng, Monash University, Australia  
 Andrew Tescher, AGT Associates (USA)  
 Jelena Tesic, IBM T.J. Watson Research Center  
 Qi Tian, University of Texas, San Antonio  
 Tsung-Han Tsai, National Central University  
 Deepak Turaga, IBM Research  
 George Tzanetakis, University of Victoria  
 Kou-Hu Tzou, ESS Technology, Inc.  
 Anthony Vetro, Mitsubishi Electric Research Laboratories, USA  
 Matthias von den Knesebeck, University of British Columbia  
 Jean-Paul Wagner, Ecole Polytechnique Federale de Lausanne  
 Chun-Hao Wang, Ryerson University  
 Haohong Wang, Qualcomm  
 Xin Wang, ContentGuard, USA  
 Ye Wang, Dept of Computer Science, National University of Singapore  
 Yong Wang, Motorola Lab, USA  
 Liu Wanquan, Curtin University  
 Bin Wei, AT&T Labs - Research  
 Zhen Wen, IBM T. J. Watson Research Center  
 Utz Westermann, University of California, Irvine  
 Lynn Wilcox, FX Palo Alto Laboratory  
 Peter Wilkins, Dublin City University, Ireland  
 Michael Wirth, University of Guelph  
 Dapeng Oliver Wu, University of Florida  
 Feng Wu, Microsoft Research Asia  
 Gang Wu, University of California, Santa Barbara  
 Guixing Wu, Dept of Electrical and Computer Engineering, Univ. of Waterloo, Canada  
 Jianhua Wu, Nanyang Technological University  
 Min Wu, University of Maryland, College Park  
 Peng Wu, Imaging Technology Department, Hewlett-Packard Labs, USA  
 Xiaolin Wu, McMaster University  
 Yi Wu, Intel Research Lab, USA  
 Yunnan Wu, Microsoft Research  
 Wang Xianbin, Communications Research Centre Canada  
 Lexing Xie, IBM T.J. Watson Research Center  
 Jun Xin, Mitsubishi Electric Research Laboratories  
 Hongkai Xiong, Shanghai Jiao Tong University of China, China  
 Yongqiang Xiong, Microsoft Research Asia  
 Changsheng Xu, Media Analysis Lab, Institute for Infocomm Research, Singapore  
 Ji-Zheng Xu, Microsoft Research Asia  
 Xiangyang Xue, Fudan University, China  
 Rong Yan, Carnegie Mellon University, USA  
 Rong Yan, IBM China Research Lab  
 Chia-Lin Yang, National Taiwan University  
 Fan Yang, Microsoft Research Asia  
 Jar-Ferr (Kevin) Yang, National Cheng Kung University, Taiwan  
 Jie Yang, Carnegie Mellon University  
 Shiqiang Yang, Tsinghua Univ.  
 Xiaokang Yang, Shanghai Jiao Tong University of China, China  
 Xiaokang Yangyang, sjtu  
 Susu Yao, Institute for Infocomm Research  
 Chia-Hung Yeh, National Dong Hwa University  
 Baocai Yin, Beijing University of Technology  
 Hao Yin, Tsinghua University  
 Naokazu Yokoya, Nara Institute of Science and Technology  
 Heather Yu, Panasonic Information and Networking Technologies Laboratory  
 Lun Yu, Fuzhou University, China  
 Rongshan Yu, Dolby Laboratories, USA  
 Wei Yu, University of Maryland College Park  
 Xiang Yu, Dept of Electrical and Computer Engineering, Univ. of Waterloo, Canada  
 André Zaccarin, Université Laval  
 Bing Zeng, Hong Kong University of Science and Technology, Hong Kong  
 Wenjun Zeng, University of Missouri - Columbia  
 Fan Zhai, Texas Instruments  
 Guangtao Zhai, SJTU  
 Cha Zhang, Microsoft Research  
 Chengcui Zhang, University of Alabama at Birmingham  
 Dengsheng Zhang, Monash University, Australia  
 Dongqing Zhang, Thomson Corporate Research, Thomson Inc.  
 Liang Zhang, Communications Research Centre Canada  
 Qian Zhang, Hong Kong University of Science and Technology, Hong Kong  
 Tong Zhang, Imaging Technology Department, Hewlett-Packard Labs, USA  
 Wenjun Zhang, Shanghai Jiao Tong University  
 Xiaoping Zhang, Ryerson University  
 Yanchao Zhang, University of Florida  
 Zhishou Zhang, Institute for Infocomm Research, Singapore  
 Zhongfei (Mark) Zhang, SUNY Binghamton  
 Debin Zhao, Institute of Computing Technology, Chinese Academy of Sciences  
 Lian Zhao, Ryerson University  
 Haifeng Zheng, Fuzhou University, China  
 Haitao Zheng, University of California, Santa Barbara  
 Sheng Zhong, Broadcom Corporation  
 Hanning Zhou, FX Palo Alto Laboratory  
 Xiang Sean Zhou, Siemens Corporation, USA  
 Bin Zhu, Microsoft Research Asia  
 Ce Zhu, Nanyang Technological University, Singapore  
 Lei Zhu, Armstrong Atlantic State University, USA  
 Xingquan Zhu, University of Vermont, USA  
 Roger Zimmerman, University of Southern California



## KEYNOTE TALKS

---

**Monday, July 10, 08:00 - 09:00, Toronto I/II**

### **Multimedia Forensics: Where Sherlock Holmes Meets Signal Processing**

*Presented by: K. J. Ray Liu*

*Chair: Ling Guan*

Dr. Liu is Professor and Associate Chair of Graduate Studies and Research, Electrical and Computer Engineering Department, University of Maryland, College Park. Dr. Liu leads the Maryland Signals and Information Group (SIG) with research contributions encompass broad aspects of information forensics and security; multimedia communications and signal processing; wireless communications and networking; biomedical imaging and bioinformatics; and signal processing algorithms and architectures, in which he has published over 400 refereed papers. He is a co-author of Multimedia Fingerprinting Forensics for Traitor Tracing, EURASIP Book Series on Signal Processing and Communication (Hindawi), 2005.

Dr. Liu is the recipient of numerous honors and awards including the National Science Foundation 1993 Research Initiation Award (RIA) and 1994 National Young Investigator (NYI) Award, the IEEE Signal Processing Society Best Paper Award in 1993 and 2005, IEEE 50th Vehicular Technology Conference Best Paper Award in 1999, EURASIP Best Paper Award in 2004, IEEE Signal Processing Society 2004 Distinguished Lecturer, and EURASIP 2004 Meritorious Service Award. Dr. Liu is a Fellow of the IEEE. He also received the 2005 Poole and Kent Company Senior Faculty Teaching Award from A. James Clark School of Engineering, University of Maryland, as well as the George Corcoran Award in 1994 for outstanding contributions to electrical engineering education, and the Outstanding Systems Engineering Faculty Award in 1996 in recognition of outstanding contributions in interdisciplinary research from Institute for Systems Research. He also received the 2004 Invention of the Year Award from University of Maryland.

Dr. Liu was the Editor-in-Chief of IEEE Signal Processing Magazine, the founding Editor-in-Chief of EURASIP Journal on Applied Signal Processing, and the prime architect and proposer of IEEE Trans. on Information Forensics and Security and IEEE Journal on Selected Topics of Signal Processing. Dr. Liu is Vice President – Publications and on the Board of Governor of IEEE Signal Processing Society. He is serving as the General Chair of IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP), Hawaii, 2007. He has served as the Chair of Multimedia Signal Processing Technical Committee, Technical Program Chair of 2003 IEEE International Conference on Multimedia and Expo, a panelist for various events of National Science Foundation, DARPA, and international conferences.

---

**Monday, July 10, 16:20 - 17:20, Toronto I/II**

### **Multimedia Signal Processing and Content-Based Image Retrieval (a Survey)**

*Presented by: A.N. Venetsanopoulos*

*Chair: Kostas N. Plataniotis*

Dr. Anastasios (Tas) N. Venetsanopoulos received the Bachelors of Engineering degree from the National Technical University in Athens (NTU), Greece, in 1965, and the M.S., P.Phil., and Ph.D. degrees in Electrical Engineering from Yale University in 1966, 1968, and 1969 respectively. He joined the Department of Electrical and Computer Engineering at the University of Toronto, Canada, in September 1968 and was appointed Inaugural Chairholder of The Bell Canada Chair in Multimedia in 1999. Between 2001-2006 he was the Dean of the Faculty of Applied Science and Engineering of the University of Toronto.

He has published over 800 papers on digital signal and image processing and digital communications and has served as Chair on numerous boards, councils and technical conference committees including IEEE committees such as the Toronto Section (1977-1979) and the IEEE Central Canada Council (1980-1982); he was President of the Canadian Society for Electrical Engineering and Vice-President of the Engineering Institute of Canada (1983-1986). He was Guest Editor and Associate Editor for numerous IEEE journals, and the Editor of the Canadian Electrical Engineering Journal (1981-1983).

He is a member of the IEEE Communications, Circuits and Systems, Computer, and Signal Processing Societies, as well as a member of Sigma Xi, the Technical Chamber of Greece, the European Association of Signal Processing, and the Association of Professional Engineers of Ontario (PEO). He was elected as a Fellow of the IEEE “for contributions to digital signal and image processing”, Fellow of EIC, “for contributions to electrical engineering”, and was awarded an Honorary Doctorate from the National Technical University of Athens, in October 1994. In October 1996, he was awarded the “Excellence in Innovation Award” of the Information Technology Research Centre of Ontario and Royal Bank of Canada, “for innovative work in colour image processing and its industrial applications”. In 2003, he was awarded the IEEE MacNaughton Medal, “for outstanding contributions to the design and implementation of communication systems, digital filters and multimedia systems; IEEE; the engineering profession and society at large”. In 2006 he was selected as the joint recipient of the 2003 IEEE Transactions On Neural Networks Outstanding Paper Award.

---

**Tuesday, July 11, 08:00 - 09:00, Toronto I/II**

## **Machine Learning for Multi-Modality Genomic Data Mining**

*Presented by: Sun-Yuan Kung*

*Chair: Hong-Jiang Zhang*

Professor S.Y. Kung received his Ph.D. Degree in Electrical Engineering from Stanford University in 1977. He was an Associate Engineer of Amdahl Corporation, Sunnyvale, 1974, and a Professor of Electrical Engineering-Systems of the University of Southern California, (1977-1987). Since 1987, he has been a Professor of Electrical Engineering at the Princeton University. He held a Visiting Professorship at the Stanford University (1984); and a Visiting Professorship at the Delft University of Technology (1984); a Toshiba Chair Professorship at the Waseda University, Japan (1984); an Honorary Professorship at the Central China University of Science and Technology (1994); and a Distinguished Chair Professorship at the Hong Kong Polytechnic University (2001-2003). His research interests include VLSI array processors, system modeling and identification, neural networks, wireless communication, sensor array processing, multimedia signal processing, bioinformatic data mining and biometric authentication.

Professor Kung is a Fellow of IEEE since 1988. He served as a Member of the Board of Governors of the IEEE Signal Processing Society (1989-1991). He was a founding member of several Technical Committees (TC) of the IEEE Signal Processing Society, including VLSI Signal Processing TC (1984), Neural Networks for Signal Processing TC (1991) and Multimedia Signal Processing TC (1998), and was appointed as the first Associate Editor in VLSI Area (1984) and later the first Associate Editor in Neural Network (1991) for the IEEE Transactions on Signal Processing. He presently serves on Technical Committees on Multimedia Signal Processing. Since 1990, he has been the Editor-In-Chief of the Journal of VLSI Signal Processing Systems.

Professor Kung has co-authored more than 400 technical publications and numerous textbooks including “VLSI and Modern Signal Processing” with Russian translation, Prentice-Hall (1985), “VLSI Array Processors”, with Russian and Chinese translations, Prentice-Hall (1988); “Digital Neural Networks”, Prentice-Hall (1993);



“Principal Component Neural Networks”, John-Wiley (1996); and “Biometric Authentication: A Machine Learning and Neural Network Approach”, Prentice-Hall (2005).

Professor Kung was a recipient of IEEE Signal Processing Society’s Technical Achievement Award for his contributions on “parallel processing and neural network algorithms for signal processing” (1992); a Distinguished Lecturer of IEEE Signal Processing Society (1994) ; a recipient of IEEE Signal Processing Society’s Best Paper Award for his publication on principal component neural networks (1996); and a recipient of the IEEE Third Millennium Medal (2000).

---

**Wednesday, July 12, 08:00 - 09:00, Toronto I/II**  
**Ten Challenging Problems in Content-Based Image And Video Retrieval**

*Presented by: Thomas Huang*

*Chair: Chang Wen Chen*

Thomas S. Huang received his B.S. Degree in Electrical Engineering from National Taiwan University, Taipei, Taiwan, China; and his M.S. and Sc.D. Degrees in Electrical Engineering from the Massachusetts Institute of Technology, Cambridge, Massachusetts. He was on the Faculty of the Department of Electrical Engineering at MIT from 1963 to 1973; and on the Faculty of the School of Electrical Engineering and Director of its Laboratory for Information and Signal Processing at Purdue University from 1973 to 1980. In 1980, he joined the University of Illinois at Urbana-Champaign, where he is now William L. Everitt Distinguished Professor of Electrical and Computer Engineering, and Research Professor at the Coordinated Science Laboratory, and Head of the Image Formation and Processing Group at the Beckman Institute for Advanced Science and Technology and Co-Chair of the Institute’s major research theme Human Computer Intelligent Interaction.

During his sabbatical leaves: Dr. Huang has worked at the MIT Lincoln Laboratory, the IBM Thomas J. Watson Research Center, and the Rheinishes Landes Museum in Bonn, West Germany, and held visiting Professor positions at the Swiss Institutes of Technology in Zurich and Lausanne, University of Hannover in West Germany, INRS-Telecommunications of the University of Quebec in Montreal, Canada and University of Tokyo, Japan. He has served as a consultant to numerous industrial firms and government agencies both in the U.S. and abroad.

Dr. Huang’s professional interests lie in the broad area of information technology, especially the transmission and processing of multidimensional signals. He has published 21 books, and over 600 papers in Network Theory, Digital Filtering, Image Processing, and Computer Vision. He is a Member of the National Academy of Engineering; a Foreign Member of the Chinese Academies of Engineering and Sciences; and a Fellow of the International Association of Pattern Recognition, IEEE, and the Optical Society of American; and has received a Guggenheim Fellowship , an A.V. Humboldt Foundation Senior U.S. Scientist Award, and a Fellowship from the Japan Association for the Promotion of Science . He received the IEEE Signal Processing Society’s Technical Achievement Award in 1987, and the Society Award in 1991. He was awarded the IEEE Third Millennium Medal in 2000. Also in 2000, he received the Honda Lifetime Achievement Award for “contributions to motion analysis”. In 2001, he received the IEEE Jack S. Kilby Medal. In 2002, he received the King-Sun Fu Prize, International Association of Pattern Recognition; and the Pan Wen-Yuan Outstanding Research Award. In 2005, he received the Okawa Prize for Information and Telecommunication Technology. In 2006, he is named “Electronic Imaging Scientist of the Year” by IS&T and SPIE.

He is a Founding Editor of the International Journal Computer Vision, Graphics, and Image Processing; and Editor of the Springer Series in Information Sciences, published by Springer Verlag.

Dr. Huang initiated the first International Picture Coding Symposium in 1969, and the first International Workshop on Very Low Bitrate Video Coding in 1993. Both meetings have become regular events (held every 12-18 months), and have contributed to the research and international standardization of image and video compression. He also (together with Peter Stucki and Sandy Pentland) initiated the International Conference on Automatic Face and Gesture Recognition in 1995. This Conference has also become a regular event, and provides a forum for researchers in this important and popular field.

## TUTORIALS

---

### **Sunday, July 9, 2006, 09:00 - 12:00**

#### **T1.1 - Steganography and Steganalysis: Fundamentals, Algorithms and Future Research**

*Presented by Prof. Yun Q. Shi, New Jersey Institute of Technology, USA*

Room: **Varley**

#### **T1.2 - Peer-to-Peer Multimedia Applications**

*Presented by Dr. Jin Li, Microsoft Research, USA*

Room: **Carmichael**

#### **T1.3 - Digital Inpainting**

*Presented by Prof. Timothy K. Shih, Tamkang University, Taiwan*

Room: **Thomson**

### **Sunday, July 9, 2006, 13:30 - 16:30**

#### **T2.1 - Capturing Light and Motion: Multimedia Techniques for Animation and Movie Production**

*Presented by Prof. Howard Leung, City University of Hong Kong and Prof. Tsuhan Chen, Carnegie Mellon University, USA*

Room: **Carmichael**

#### **T2.2 - Hardware and Software Techniques for Multimedia Systems**

*Presented by Prof. Nikil Dutt, University of California at Irvine, USA and Prof. Wayne Wolf, Princeton University, USA*

Room: **Thomson**

#### **T2.3 - Music Information Retrieval**

*Presented by Prof. George Tzanetakis, University of Victoria, Canada*

Room: **Varley**

## REGISTRATION

---

Registration will be in the conference area foyer and will be available at the following times:

Sun, July 9 ..... 08:30 - 17:00  
Mon, July 10 ..... 07:00 - 17:30  
Tue, July 11 ..... 08:00 - 16:00  
Wed, July 12 ..... 08:00 - 16:00

## EXHIBITION

---

ICME 2006 would like to thank our partners who are exhibiting. Please take time to visit these exhibitors and express your appreciation for their participation in ICME 2006.

Epson Canada  
National Chiao Tung University/Academia Sinica Taiwan  
ATI Technologies Inc.  
John Wiley & Sons, Ltd.  
Springer

Exhibits will be open according to the following schedule:

Mon, July 10 ..... 09:00 - 16:30  
Tue, July 11 ..... 09:00 - 16:00  
Wed, July 12 ..... 09:00 - 16:30

## MEETING SCHEDULE

---

### **Mon, July 10**

12:00 - 14:00

- CAS Society Multimedia Systems TC Meeting — Room: Richmond
- SP Society Multimedia Signal Processing TC Meeting — Room: Osgoode

17:30 - 22:30

- MM Transactions Steering Committee Meeting — Room: Governor General Suite

### **Tue, July 11**

12:00 - 14:00

- MM Transactions on Multimedia Editorial Board Meeting — Room: Richmond

### **Wed, July 12**

12:00 - 14:00

- ICME Steering Committee Meeting — Room: Richmond

# WELCOME TO TORONTO

---

Welcome to Toronto, the largest city in Canada and the fifth largest city in North America. Toronto is a city with a long history of multiculturalism and growth, tempered with a sense of community and tradition that has allowed Toronto to become one of the greatest cities in the world, with strong influences from many sources melding into a unique whole. Situated on the shores of Lake Ontario, Toronto is farther south than many American cities, but it's nature is distinctly Canadian despite its proximity to the States. An amalgam of six smaller cities that originally became affiliated in 1953 to ease population growth after the second World War, in 1998 the individual sections of Toronto were officially combined into what is today called the city of Toronto.

Toronto is also a strong force in the arts, whether musical or visual, live or recorded. With a progressive spirit and a sense of its identity and promising future, Toronto is quickly becoming one of the best known and loved cities in the world today.

## CLIMATE IN TORONTO

---

July tends to be the middle of the summer season in Toronto. The average daily high temperatures for Toronto in July are 84°F or 29°C, with average low temperatures of 64°F or 17°C.

## ENTRY REQUIREMENTS

---

### **TO VISIT CANADA YOU:**

- \* must be healthy. You might need a doctor's examination;
- \* must respect Canadian laws;
- \* will need a valid passport, proof of who you are or other travel documents; and
- \* may need a Temporary Resident Visa (TRV)\*\*.

Visit <http://www.cic.gc.ca/english/visit/visas.html> to find the list of who will need a VISA. Citizens of U.S. and many European countries need no visa, but visit the above website for a more comprehensive list.

<http://www.cic.gc.ca/english/applications/visa.html> contains information on how to get your VISA.

The VISA application process should be initiated well in advance of your estimated departure date.

Canada does not pay for hospital or medical services for visitors. Make sure you have health insurance to pay your medical costs before you leave for Canada.

Contact a Canadian embassy, high commission or consulate for information on what you will need before coming to Canada. Most ICME attendees will not need a VISA, but they will need a passport.

### **ONCE YOU ARRIVE:**

A custom's officer will ask you a few short questions when you arrive. To make this go quickly, keep your passport with you and not in your luggage.

The officer will stamp your passport or advise you how long you may stay in Canada. Feel free to ask questions if you are unsure about anything.

For VISAS, most applicants must have a passport and one passport size photograph. It may be helpful for an applicant to have a letter of invitation from the conference. If you need a personal letter of invitation to attend ICME 2006, please contact Mrs. Billene Mercer, Conference Management Services, [mercerc@cmsworldwide.com](mailto:mercerc@cmsworldwide.com). You should provide her with your complete mailing address, e-mail address, and fax number. The letter of invitation with an original signature will be sent to you immediately.

---

## **CUSTOMS INFORMATION**

---

The website for general customs information is: <http://www.ccra-adrc.gc.ca> and <http://www.ccra-adrc.gc.ca/customs/individuals/menu-e.html>.

---

## **TORONTO AIRPORTS**

---

### **LESTER B. PEARSON INTERNATIONAL AIRPORT**

Lester B. Pearson International Airport (also known as LBPIA) is 16 miles (27 km) northwest of downtown Toronto. Travel time from the downtown area to the airport is generally 30-40 minutes.

You may want to check the Internet for discount airfares at: [www.airfare.com](http://www.airfare.com). Other reliable consolidators to consult for low fares include [www.counciltravel.com](http://www.counciltravel.com) or [www.statravel.com](http://www.statravel.com). Non-U.S. and non-Canadian travelers can take advantage of the Visit USA airpasses offered by Continental (800 231-0856) and similar coupons from Air Canada (800 776-3000).

The VIA Rail network is Canada's train system that can also be used to enter Toronto from other parts of Canada. All long distance trains arrive at Union Station, which is directly linked to the Toronto subway system. Further information can be retrieved at: (888) 842-7425 or [www.viarail.ca](http://www.viarail.ca). Toronto can also be reached by train using the GO Transit commuter train system, (416) 594-1010 or Amtrak, (800) 872-7245.

The Toronto Airport Express is a regular bus service available that runs from the airport into Toronto and the surrounding districts. A bus ticket costs CN\$16.75. More information can be found by calling (800) 387-6787 or by visiting the bus service website at [www.torontoairportexpress.com](http://www.torontoairportexpress.com).

Taxi services between Pearson International Airport and Toronto cost between CN\$40 and CN\$50. Taxis are located outside all terminals and are metered.

### **TORONTO CITY CENTRE AIRPORT**

Attendees may also choose to fly into Toronto through the smaller Toronto City Centre Airport (TCCA), also known as the Toronto Island Airport. This smaller airport is only serviced by Air Canada airlines, however, so may only be suitable to attendees coming from nearby Canada or the United States. The TCCA is located close to the downtown area on the western side of the Toronto Islands, and a passenger and vehicle ferry travels across the 400 ft (121 m) wide Western Channel to Toronto every 15 minutes while the airport is in operation.

Greyhound Lines also provide transportation from certain U.S. cities into Toronto. Greyhound Canada (800-661-8747).

**DEPARTURE TAXES**

There is an airport departure tax of \$55 placed on all international flights out of Canada, except those to US destinations. For the US destinations, the tax is 7% of the ticket value plus \$6 to a maximum of \$55. Most tickets include the departure tax; however, if your ticket was purchased outside of Canada, it may not include this tax.

For international flights, you should check in two hours before the scheduled departure.

**TAXI SERVICE**

Taxi service in Toronto is generally reliable. An excellent taxi service that provides both standard and handicap-accessible taxis is Beck Taxi at (416) 751-5555.

**RENTAL CAR INFORMATION**

---

There are many car rental companies in Toronto. Below is a list of major agencies. Check the telephone directory for additional companies. If you wish to rent a car and are not a Canadian or United States citizen, you must obtain an International Driving Permit.

Avis.....	(514) 636-1902 (800) 879-2847 in Canada or (800) 331-1084
Budget.....	(800) 268-8900 or (514) 636-0052
Discount.....	(514) 286-1554 or (800) 263-2355
Dollar .....	(514) 344-5858 downtown or (800) 800-6000
Enterprise.....	(800) 562-2886 or (514) 931-3722
Hertz .....	(800) 263-0600 in Canada or (800) 654-3001 or (514) 842-8537
National Tilden.....	(514) 878-2771 or (800) 387-4747

**CURRENCY**

---

The Canadian dollar (\$) is divided into 100 cents (¢). Coins come in denominations of 1¢ (penny), 5¢ (nickel), 10¢ (dime), 25¢ (quarter), 1\$ (loonie) and 2\$ (twoonie) pieces. Bills come in \$5, \$10, \$20, \$50 and \$100 denominations. \$50 and \$100 dollar bills are not accepted by some shops or services.

You may check the Currency Exchange Services at: [www.foreignmoney.com](http://www.foreignmoney.com) or from [www.us.thomascook.com](http://www.us.thomascook.com).

ATMS are found in most of the same places as in the United States.

Most major credit cards are accepted. Visa and MasterCard dominate the market, followed by American Express, Diners Club, and enRoute.

You can also exchange your cash at the airport or at major banks downtown. Banks and other financial institutions offer a standard rate of exchange; the best exchange rates can be obtained by withdrawing funds from bank's ATMs.

American Express.....(800) 807-6233  
Thomas Cook.....(800) 287-7362

## **TAXATION**

---

Most goods and services in Canada are taxed 7% by the federal government and appears on the bill as the TPS tax or also known as GST. In addition to this tax, Quebec also charges an 8% provincial tax on goods and services, including hotel accommodations. This tax appears as the PST. The total tax on most goods and services is thus 15%. Visitors to Toronto from outside the country can apply to have both the federal and provincial tax returned to them upon their departure. Be sure to keep all purchase receipts as proof of payment. This application form can be found in the Tax Refund for Visitors to Canada booklet. It's available at most hotels, Tourist Information Centers, travel agents, duty-free shops, and shopping centers. You can also call the Visitor Rebate Program (800) 668-4748 (within Canada) or (902) 432-5608 (outside Canada); or visit the website: [www.ccr-a-adrc.gc.ca](http://www.ccr-a-adrc.gc.ca).

## **GETTING AROUND**

---

### **PUBLIC TRANSPORTATION**

The subway system is an efficient and safe way to get around Toronto. The subway operates from Monday to Friday 6 am to 1:30 am, and Sunday 9 am to 1:30 am. Outside of these hours, a surface route known as the Blue Night Network runs roughly every 30 minutes. The subway system consists of a north-south line (Yonge-University-Spadina) and an east-west line (Bloor-Danforth). The system is interconnected with the bus and streetcar network, and transfer tickets can be picked up at the stations or from a bus driver, which is cheaper than buying separate tickets for each mode of transportation. Free system maps (also known as ride guides) can be obtained at any subway station to provide complete route information. For more information about the subway system call (416) 393-4636 between 7 am and 10 pm Eastern Standard Time.

## CONFERENCE VENUE & HOW TO GET THERE

---

ICME 2006 will be held in the award-winning Hilton Toronto, located near the center of the downtown area. The hotel information is as follows:

Hilton Toronto  
145 Richmond Street West  
Toronto, Ontario, Canada M5H 2L2  
(416) 869-3187

### **General directions to the Hilton Toronto:**

**From 401 West:** Follow Hwy 401 eastbound to Hwy 427. Take Highway 427 south to the Gardiner Expressway east. Exit onto York Street, head North to Richmond Street, turn left, (one way heading west) and the Hilton Toronto is located on the left-hand side.

**From 401 East:** Follow Hwy 401 Westbound to the Don Valley Parkway (DVP). Take DVP south to Richmond Street Exit. (Richmond Street is one way) Follow Richmond Street Westbound. After Bay Street, get into the left hand lane. Hilton Toronto is the second building on the left hand side after York Street.

**From QEW West:** Take the QEW (Queen Elizabeth Way) to Toronto. It becomes the Gardiner Expressway. Follow the Gardiner Expressway to the York Street exit. Take York Street north to Richmond Street West. Turn left on Richmond Street. The Hilton Toronto is the second building on the left.

## NEIGHBORHOODS

---

### **TORONTO ISLANDS**

These islands, which were originally a peninsula until a storm separated them in 1858, cover approximately 600 acres south of Toronto's harbor. They once held opulent hotels and retreats for high society, but today they are home to shady biking and hiking trails, tennis courts, picnic sites and an old-fashioned amusement park complete with a carousel. The islands are a 20-minute ferry ride from the city, and are known to give a spectacular view of the Toronto skyline. For those seeking more exercise, bikes can be rented onsite.

### **YORKVILLE**

*Yorkville Ave. Between Yonge St. & Avenue Rd.*

Although this neighborhood was once showing its age, recent renovations and investments have transformed it into one of the most fashionable and chic areas of the city. The Victorian houses that line the streets have been converted into cafes, salons and shops catering to upscale patrons. These tree-lined avenues provide a perfect place for a quiet afternoon of shopping or simply enjoying the city.

### **CABBAGETOWN**

*East of Parliament St., between Wellesley & Dundas St. East*

Originally a poor immigrant neighborhood, Cabbagetown's name is thought to come from the many cabbages sprouting in the front-yard gardens of the area. Today, however, Cabbagetown is a network of green spaces,



trees and renovated homes. Due to the ethnically diverse background of the area, restaurants include such fare as Ethiopian, Chinese, French, Japanese and more. The area also includes the 19th century Riverdale Farm, Allan Gardens, and many colorful shops.

## **CHINATOWN**

*Dundas St. from Elizabeth St. to Spadina Ave.*

Toronto's Chinatown is the center of one of the largest Chinese districts in North America. Street vendors, especially produce sellers, fill the streets and byways of the area. Asian foods and goods are sold from many stores and restaurants, and the smells of cooking and flower shops fill the streets. This is an excellent neighborhood for a day visit or a stroll through the city.

## **GREEKTOWN**

*Danforth Ave., between Chester St. & Broadview Ave.*

[www.greektowntoronto.com](http://www.greektowntoronto.com)

Similar to Chinatown in some ways, the streets of Greektown also bustle with vendors and fruit markets. Trendy shops and cafes line the avenues festooned with the Greek flag. This family-friendly area is known for its welcoming nature, pastry shops, and memorable sights.

## **KENSINGTON MARKET**

*Kensington Ave., west of Spadina Ave. and north of Dundas St.*

What was once a Jewish marketplace has become a home for the Portuguese, Caribbean and East Indian communities of Toronto, creating an atmosphere like no other. Everything from imported food boutiques to vintage clothing shops to fruit and vegetable stands can be found along the streets of Kensington Market.

## **LITTLE INDIA**

*Gerrard St. between Woodfield & Coxwell*

[www.gerrardindianbazaar.com](http://www.gerrardindianbazaar.com)

Also known as the Gerrard Indian Bazaar, this stretch of Gerrard St. contains over one hundred Indian, Pakistani, Sri Lankan and Bangladeshi stores. Food markets and jewelry shops abound, but the area is known for its fabric stores and clothing shops featuring saris and other traditional fashions. Over 50 Indian restaurants occupy the Bazaar, and many of them feature live performances of classical Indian music.

## **LITTLE ITALY**

*College St. from Euclid Ave. to Shaw St.*

This Italian district has recently become one of the city's nighttime hot spots, with a variety of trendy bars and restaurants. The area is dense with restaurants and bistros, and often full of revelers. For a quieter urban-Italian experience, the smaller neighborhood of Corso Italia may be more appropriate. (St. Clair Ave., west of Bathurst St. To Lansdowne Ave.) Slightly to the north of Little Italy, this region is a more "traditional" Italian neighborhood, interspersed with family-run restaurants and cafes.

## **THE BEACHES**

*Eastern Queen St.*

This quiet community has a small-town feel that makes it a favorite of baby-boomers, whether they are visiting or local. Quirky shops and antique stores combined with a boardwalk bordering a lake make this an excellent area for a relaxed day trip inside the city.

## **QUEEN STREET WEST**

*Western Queen St.*

The home and headquarters of the hip and fashionable, Queen Street West is a favorite of students, social climbers and bohemians alike. Clothing designers and fabric outlets rub shoulders here with second-hand bookstores and the cutting edge of the Toronto music scene. This funky neighborhood is sure to be a favorite of those who like the young nightlife.

## **PLACES TO EXPLORE**

---

Every corner of Toronto is filled with exciting and memorable things to explore and experience. Below are listed just a few of the things that make Toronto popular with tourists from around the world. For more suggestions, visit [www.torontotourism.com](http://www.torontotourism.com).

### **CN TOWER**

The CN Tower is the tallest free-standing structure in the world, measuring 180 stories (553 m/1815 ft) at its tip. The glass floored areas, turbo elevator and revolving restaurant make visiting the Tower and unforgettable experience.

### **ROY THOMSON HALL**

This concert hall has been designed to provide the best possible acoustic environment for anything that is performed there. Outfitted with the most sophisticated acoustic technologies and an interior shape that can change to suit different performance styles, the Roy Thomson Hall is fascinating to visit even when there is no performance scheduled.

### **ROYAL ONTARIO MUSEUM**

The Royal Ontario Museum contains so many treasures that it would be impossible to see them all in a single visit. From ancient Egyptian antiquities and European suits of armor to dinosaur skeletons and totem poles, this museum holds something in its four levels and fifty exhibit spaces for any visitor.

### **ART GALLERY OF ONTARIO**

On the edge of Chinatown, the Art Gallery of Ontario contains over 26,000 works of art ranging from 15th century European paintings to contemporary sculpture and modern Inuit carvings. Anyone who enjoys visual art of any kind is sure to find something enjoyable and interesting within the AGO's spacious halls and galleries.

## **HOCKEY HALL OF FAME**

The Stanley Cup, jerseys of famous players, videos of the game's greatest moments and even a reproduction of a dressing room for the stars are all here at the Hockey Hall of Fame. The Hall has been open for over thirty years, and is a must for any visiting hockey fan.

## **BLACK CREEK PIONEER VILLAGE**

Since it first opened in 1960, Black Creek Pioneer Village has been recreating life in a 19th century farming community for all who want to see it. A tinsmith, a printing press and a waterwheel mill are some of the various sights to see in this quaint and charming recreation.

## **TOMMY THOMPSON PARK**

Tommy Thompson Park, which is located on a eight mile long peninsula into Lake Ontario, has been designated an Important Bird Area (IBA), and is home to several large bird populations, including terns, cormorants, inland gulls and herons. Other wildlife, such as rabbits, coyotes, woodchucks and butterflies also make the park their home. Visitors are free to cycle and walk on the trails, but dogs and other pets are not allowed.

## **METRO TORONTO ZOO**

Containing more than 5,000 animals, six "zoogeographic regions" and 710 acres, the Metro Toronto Zoo is a world class attraction to animal lovers and the young at heart. The zoo is home to such rare species as the Indian rhinoceros, the pygmy hippopotamus and the lowland gorilla, not to mention Canada's largest herd of African elephants. The zoo also features a "Zoomobile" which gives tours of the zoo as a whole.

## **DAY CRUISES**

Toronto's panoramic skyline becomes even more dramatic when it is seen from the water. Cruises ranging from a two-hour harbor tour to an all-day cruise with a shipboard dinner and dancing can be scheduled, on ships as diverse as enclosed tour boats, three-masted Great Lakes schooners ([www.greatlakesschooner.com](http://www.greatlakesschooner.com)), or even Canada's tallest sailing ship, the Empire Sandy ([www.empiresandy.com](http://www.empiresandy.com)). For schedules and fares, check [www.harbourfront.on.ca](http://www.harbourfront.on.ca).

## **HIPPO TOURS**

These amphibious bus tours seat forty and take passengers on a ninety-minute journey around Toronto, cruising through both city streets and harbor waves. The tours leave every hour, on the hour. More information can be found at [www.torontohippotours.com](http://www.torontohippotours.com).

---

## **DAY TRIPS**

---

### **NIAGARA FALLS**

Roughly 80 miles (130 kilometers) from Toronto, the city of Niagara Falls is a relaxing destination for a day trip away from the big city. Boasting more than just its namesake waterfalls, the "Honeymoon Capital of the World" has many well maintained museums and parks, as well as shopping casinos and the aquatic park "Marineland". Of course, no visit to Niagara would be complete without seeing the magnificent falls, especially from the deck of

the tour ship Maid of the Mist. Tours can be arranged through [www.niagarafallstours.net](http://www.niagarafallstours.net), and more information on the city can be found at the city’s webpage, [www.city.niagarafalls.on.ca](http://www.city.niagarafalls.on.ca).

**AFRICAN LION SAFARI**

If the Toronto nightlife isn’t wild enough for you, the African Lion Safari ([www.lionsafari.com](http://www.lionsafari.com)) is only about an hour from the city, and is sure to thrill the entire family. The park houses more than 1000 animals, including lions, elephants, baboons, zebra and more, on a series of 5 to 50 acre (2 to 20 hectare) reserves. Guests can either drive through in their own car or join a guided tour, and there are many animal performances to enjoy as well. Fair warning to all tourists: with the petting zoos, breeding programs and boat and railway tours, you may need more than one trip to take it all in.

**TIPPING**

---

Tipping is generally not included in service charges in Canada. The general rule is to tip 15% of the entire bill. Porters and doormen should get about \$2 a bag and maids should receive about \$2 per person per day (\$3 in luxury hotels).

**HELPFUL PHONE NUMBERS**

---

- Tourist Information..... (416) 203-2600
- Police (non-emergency)..... (416) 808-2222
- Poison Information ..... (416) 813-5900
- All-Night Pharmacies:
  - 3089 Duffering St. at Lawrence St. .... (416) 787-0238
  - 2345 Yonge St. at Eglinton St. .... (416) 487-5411
  - 700 Bay St. at Gerard St. .... (416) 979-2424
- Emergencies:
  - Police, fire, and ambulance ..... 911
  - Dental Emergency Service (8am-midnight) ..... (416) 485-7121

**MA1-L1 Feature Extraction and Segmentation I (Lecture)**

Time: Monday, July 10, 09:30 - 10:50

Place: Carmichael

Chair: Kiyoharu Aizawa, University of Tokyo

09:30 - 09:50

**MA1-L1.1 KEY FRAME EXTRACTION IN 3D VIDEO BY RATE-DISTORTION OPTIMIZATION**

*Jianfeng Xu, Toshihiko Yamasaki, Kiyoharu Aizawa, University of Tokyo, Japan*

3D video, which consists of a sequence of 3D mesh models, can provide detailed 3D information both in spatial and temporal domain. In this paper, a key frame extraction method has been developed to summarize 3D video by rate-distortion (R-D) optimization. For this purpose, we introduce an effective feature vector extraction algorithm from 3D video. Prior to key frame extraction, shot detection is performed using the feature vectors as a pre-processing. Then, a rate-distortion curve is generated in each shot, where the locations of key frames are optimized. Lastly, R-D trade-off can be achieved by optimizing a cost function with a Lagrange multiplier. Our experimental results show the extracted key frames are compact and faithful to original 3D video.

09:50 - 10:10

**MA1-L1.2 EVOLUTIONARY FEATURE GENERATION IN SPEECH EMOTION RECOGNITION**

*Bjoern Schuller, Stephan Reiter, Gerhard Rigoll, Technische Universität München, Germany*

Feature sets are broadly discussed within speech emotion recognition by acoustic analysis. While popular filter and wrapper based search help to retrieve relevant ones, we feel that automatic generation of such allows for more flexibility throughout search. The basis is formed by dynamic Low-Level Descriptors considering intonation, intensity, formants, spectral information and others. Next, systematic derivation of prosodic, articulatory, and voice quality high level functionals is performed by descriptive statistical analysis. From here on feature alterations are automatically fulfilled, to find an optimal representation within feature space in view of a target classifier. To avoid NP-hard exhaustive search, we suggest use of evolutionary programming. Significant overall performance improvement over former works can be reported on two public databases.

10:10 - 10:30

**MA1-L1.3 IDENTIFICATION AND DETECTION OF THE SAME SCENE BASED ON FLASHLIGHT PATTERNS**

*Masao Takimoto, Tokyo University, Japan; Shin'ichi Satoh, Masao Sakauchi, National Institute of Informatics, Japan*

A method has been developed to identify video shots of the same scene where camera flash lights are observed, and the method has been tested by using it to detect such shots from a large TV video archive. Camera flashes are often used in impressive scenes, such as interviews of important persons. Because such scenes are broadcasted repeatedly on various TV programs, a method for detecting them is a promising approach for semantic video indexing. The proposed identification method is invariant to the differences in viewpoint, illumination or any other visual environment because it depends on comparison between temporal occurrence patterns of flash lights. Furthermore, because each flash pattern is represented with a binary array, the comparison requires low computing cost. These advantages mean that the proposed method can be considered to provide semantic and efficient video analysis.

10:30 - 10:50

**MA1-L1.4 DETECTING MUSICAL SOUNDS IN BROADCAST AUDIO BASED ON PITCH TUNING ANALYSIS**

*Yongwei Zhu, Qibin Sun, Susanto Rahardja, Institute for Infocomm Research, Singapore*

Detecting the presence of musical sounds in broadcast audio is important for content-based indexing and retrieval of auditory and visual information in radio and TV programs. In this paper, we propose a novel approach for musical sounds detection in broadcast audio based on the analysis of the characteristic feature of musical tones, pitch tuning. A spectral analysis method is presented for detecting the evidence of pitch tuning in the audio signal. Unlike the existing methods for discriminating speech and music, the proposed technique is not limited by inadequate training data, and it can deal with the case of music mixed with speech. In addition, the technique can be efficiently implemented for real-time application. Experiments based on TRECVID data set have shown good performance of the proposed technique.

**MA1-L2 Processors & Multimedia I (Special Session)**

Time: Monday, July 10, 09:30 - 10:50

Place: Johnston

Co-Chairs: Yen-Kuang Chen, Intel Corporation and Eric Debes, Intel Corporation

09:30 - 09:50

**MA1-L2.1 SCALABILITY OF MULTIMEDIA APPLICATIONS ON NEXT-GENERATION PROCESSORS***Guy Amit, Yaron Caspi, Ran Vitale, Adi Pinhas, Intel Corporation, Israel*

In the near future, the majority of personal computers are expected to have several processing units. This is referred to as Core Multiprocessing (CMP). Furthermore, each of the computation units will be capable of running multiple hardware threads. To benefit from the additional processing power, application developers should multithread their software. This paper studies the scalability (expected speedup factor) of multimedia applications and provides guidelines for proper utilization of these new multi-core platforms. In particular, we discuss the decomposition method, load balancing, synchronization primitives, interaction with the operating system and hardware issues such as cache hierarchy and memory bandwidth. Our results are based on analysis of several state-of-the-art applications, including H.264 video encoding, panoramic image stitching and dense optical-flow estimation. We demonstrate how to multithread them properly, and report scalability results on several next-generation multi-core platforms.

09:50 - 10:10

**MA1-L2.2 ON PARALLELIZATION OF A VIDEO MINING SYSTEM***Wenlong Li, Eric Li, Nan Di, Carole Dulong, Tao Wang, Yimin Zhang, Intel Corporation, China*

As digital video data becomes more pervasive, mining information from multimedia data becomes increasingly important. Although researches in multimedia mining area have shown great potential in daily life, the huge computational requirement prohibits its wide use in practice. Since our personal computer is shifting from uniprocessors to multicore processors, exploiting thread level parallelism in multimedia mining applications is critical to utilize the hardware resources and accelerate the mining process. This paper presents three different parallel approaches (task level, data slicing and hybrid parallel) to parallelize one widely used application in video mining system. The hybrid scheme, with the exploration of data level and task level parallelism, delivers much better performance than other two schemes. We get 10x performance improvement on a 16-way multiprocessor system. Besides, we perform several efficient optimization techniques, such as subexpression optimization, SIMD, and data blocking, to improve the performance by more than 60%. Therefore, our parallelization and optimization of the application makes it 16x faster than it used to be. Our study shows that with proper parallelization and optimization, multimedia mining can be used widely in our daily life soon.

10:10 - 10:30

**MA1-L2.3 COLLABORATIVE MULTITHREADING: AN OPEN SCALABLE PROCESSOR ARCHITECTURE FOR EMBEDDED MULTIMEDIA APPLICATIONS***Wei-Chun Ku, Shu-Hsuan Chou, Jui-Chin Chu, Chih-Heng Kang, Tien-Fu Chen, Jiun-In Guo, National Chung Cheng University, Taiwan*

Numerous approaches can be employed in exploiting computation power in processors such as superscalar, VLIW, SMT and multi-core on chip. In this paper, a VisoMT processor is proposed, which unified VLIW and multithreading by providing an efficient control and data communication model, while offering explicit parallelisms for embedded applications. The architecture concurrently executes a main thread and several accelerative threads, coordinated by the main thread. A switch-based register-file is provided for fast data exchange between these accelerative threads. Moreover, a SMT helper function unit is introduced for program controls of accelerative threads, and an event-driven mechanism is introduced for synchronization between the main thread and these accelerative threads. In the final, we will show the area and speed of each module and the data of chip.

10:30 - 10:50

**MA1-L2.4 VIDEO ANALYSIS AND COMPRESSION ON THE STI CELL BROADBAND ENGINE PROCESSOR**

*Lurng-Kuo Liu, Sreeni Kesavarapu, Jonathan Connell, Ashish Jagmohan, Lark-hoon Leem, Brent Paulovicks, Vadim Sheinin, Lijung Tang, Hangu Yeo, IBM T. J. Watson Research Center, United States*

With increased concern for physical security, video surveillance is becoming an important business area. Similar camera-based system can also be used in such diverse applications as retail-store shopper motion analysis and casino behavioral policy monitoring. There are two aspects of video surveillance that require significant computing power: image analysis for detecting objects, and video compression for digital storage. The new STI CELL Broadband Engine (CBE) processor is an appealing platform for such applications because it incorporates 8 separate high-speed processing cores with an aggregate performance of 256Gflops. Moreover, this chip is the heart of the new Sony Playstation 3 and can be expected to be relatively inexpensive due to the high volume of production. In this paper we show how object detection and compression can be implemented on the CBE, discuss the difficulties encountered in porting the code, and provide performance results demonstrating significant speed-up.

**MA1-L3: Monday, July 10, 09:30 - 10:50**

**MA1-L3 Content Understanding I (Lecture)**

Time: Monday, July 10, 09:30 - 10:50

Place: Thomson

Chair: Wei-Ying Ma, Microsoft Research Asia

09:30 - 09:50

**MA1-L3.1 COMPUTATIONAL INTELLIGENCE TECHNIQUES AND THEIR APPLICATIONS IN CONTENT-BASED IMAGE RETRIEVAL**

*Kambiz Jarrah, Matthew Kyan, Sri Krishnan, Ling Guan, Ryerson University, Canada*

The main focus of this paper is to present a methodology for optimizing relevance identification in content-based image retrieval (CBIR) systems through the principle of feature weight detection. The purpose of relevance identification is to find a collection of images that are statistically similar to, or match with, an original query image within a large visual database. The novelty of this scheme is two-fold: using a base-10 Genetic Algorithm method to accurately determine the contribution of individual feature vectors for a successful retrieval in the so-called feature weight detection process, and defining a new unsupervised learning algorithm, the Directed self-organizing tree map (DSOTM), for the purpose of classification in the automatic relevance identification module of the search engine. Comprehensive experiments demonstrate feasibility of the proposed methodology.

09:50 - 10:10

**MA1-L3.2 VIDEO NEWS SHOT LABELING REFINEMENT VIA SHOT RHYTHM MODELS**

*John R. Kender, Columbia University, United States; Milind R. Naphade, IBM T. J. Watson Research Center, United States*

We present a three-step post-processing method for increasing the precision of video shot labels in the domain of television news. First, we demonstrate that news shot sequences can be characterized by rhythms of alternation (due to dialogue), repetition (due to persistent background settings), or both. Thus a temporal model is necessarily third-order Markov. Second, we demonstrate that the output of feature detectors derived from machine learning methods (in particular, from SVMs) can be converted into probabilities in a more effective way than two suggested existing methods. This is particularly true when detectors are errorful due to sparse training sets, as is common in this domain. Third, we demonstrate that a straightforward application of the Viterbi algorithm on a third-order FSM, constructed from observed transition probabilities and converted feature detector outputs, can refine feature label precision at little cost. We show that on a test corpus of TREC-VID 2005 news videos annotated with 39 LSCOM-lite features, the mean increase in the measure of Average Precision (AP) was 4%, with some of the rarer and more difficult features having relative increases in AP of as much as 67%.

10:10 - 10:30

**MA1-L3.3 WHICH THOUSAND WORDS ARE WORTH A PICTURE? EXPERIMENTS ON VIDEO RETRIEVAL USING A THOUSAND CONCEPTS**

*Wei-Hao Lin, Alexander Hauptmann, Carnegie Mellon University, United States*

In contrast to traditional video retrieval that represents visual content with low-level features (e.g. color and texture), emerging concept-based video retrieval allows users to search video archives by specifying a limited number of high-level concepts (e.g. outdoors and car). Recent studies have demonstrated the feasibility of concept-based retrieval, but a fundamental question remains: what kinds of concepts should be indexed? By analyzing a large video archive annotated with more than a thousand high-level concepts we develop guidelines for choosing concepts of high utility in video retrieval.

10:30 - 10:50

**MA1-L3.4 CLUSTERING-BASED ANALYSIS OF SEMANTIC CONCEPT MODELS FOR VIDEO SHOTS**

*Markus Koskela, Alan Smeaton, Dublin City University, Ireland*

In this paper we present a clustering-based method for representing semantic concepts on multimodal low-level feature spaces and study the evaluation of the goodness of such models with entropy-based methods. As different semantic concepts in video are most accurately represented with different features and modalities, we utilize the relative model-wise confidence values of the feature extraction techniques in weighting them automatically. The method also provides a natural way of measuring the similarity of different concepts in a multimedia lexicon. The experiments of the paper are conducted using the development set of the TRECVID 2005 corpus together with a common annotation for 39 semantic concepts.

**MA1-L4: Monday, July 10, 09:30 - 10:50**

**MA1-L4 Transcoding X to H.264 I (Special Session)**

Time: Monday, July 10, 09:30 - 10:50

Place: Toronto I

Chair: Hari Kalva, Florida Atlantic University

09:30 - 09:50

**MA1-L4.1 OVERVIEW OF LOW-COMPLEXITY VIDEO TRANSCODING FROM H.263 TO H.264**

*Jens Bialkowski, Marcus Barkowsky, André Kaup, University of Erlangen-Nuremberg, Germany*

H.264 standardized by the ITU-T and ISO/IEC is the latest video standard with the highest compression efficiency. It will be used in future multimedia devices such as handhelds using DVB-H or mobile phones. Interoperability between new and old applications is very important. Thus it will be necessary to transcode existing video bitstreams to H.264. Because of the many new prediction parameters as well as the pixel-based deblocking filter and the new transform of H.264 this is a difficult task to perform. In our work we propose a fast cascaded pixel-domain transcoder from H.263 to H.264 for both intra- and inter-frame coding. The rate-distortion performance of the encoded bitstreams is compared to an exhaustive full-search approach. Our approach leads to less than 10% higher data rate, but the computational complexity for the prediction can be reduced by 90% and more. It will be shown that the algorithms proposed for H.263 are applicable for transcoding MPEG-2 to H.264, too.



09:50 - 10:10

**MA1-L4.2 AN EFFICIENT ALGORITHM FOR VC-1 TO H.264 VIDEO TRANSCODING IN PROGRESSIVE COMPRESSION**

*Jae-Beom Lee, Sarnoff Corporation, United States; Hari Kalva, Florida Atlantic University, United States*

The high definition video adoption has been growing rapidly for the last two years. The two high definition DVD formats HD-DVD and Blu-ray have mandated MPEG-2, H.264 and VC-1 as video compression formats. The coexistence of these different video coding standards creates a need for transcoding. In this paper, an efficient transcoding algorithm from VC-1 video to H.264 video is discussed. While there has been recent work on MPEG-2 to H.264 transcoding, the published work on VC-1 to H.264 transcoding is non-existent. There is very limited amount of published work on VC-1. This paper gives a brief overview of VC-1 and discusses the opportunities for low-complexity tools for VC-1 to H.264 transcoding. The paper considers I and P frame transcoding from VC-1 to H.264 and proposes two approaches 1) Low Cost Design and 2) High Cost Design.

10:10 - 10:30

**MA1-L4.3 MPEG-4 TO H.264 TRANSCODING USING MACROBLOCK STATISTICS**

*Yung-Ki Lee, Sung-Sun Lee, Yung-Lyul Lee, Sejong University, Republic of Korea*

In this paper, a temporal resolution reduction transcoding method that transforms an MPEG-4 video bitstream into an H.264 video bitstream is proposed. The block mode statistics and motion vectors in the MPEG-4 bitstream are utilized in the H.264 encoder for block mode conversion and motion vector interpolation methods. The proposed motion vector interpolation methods are developed not to perform brute-force motion estimation again in the H.264. In the experimental results, the proposed methods achieve 3~4 times improvement in computational complexity compared to the cascade pixel-domain transcoding, while the PSNR (peak signal to noise ratio) is degraded with 0.2~0.9dB depending on the bitrates.

10:30 - 10:50

**MA1-L4.4 A FAST INTRA MODE DECISION ALGORITHM FOR AVS TO H.264 TRANSCODING**

*Zhihang Wang, Wen Gao, Chinese Academy of Sciences, China; Debin Zhao, Harbin Institute of Technology, China; Qingming Huang, Chinese Academy of Sciences, China*

H.264/AVC is the latest video coding standard developed by MPEG and ITU. AVS is the latest video coding standard developed by China. The two new video coding standards will co-exist in the market in the future. So it is necessary to transcode the AVS format to the H.264/AVC format or vice versa. Because the AVS video coding standard is based on the 8x8 block and also uses the intra prediction technology, it is possible to use the AVS intra mode information to predict the H.264/AVC intra mode. In this paper, a fast intra mode decision algorithm was proposed to reduce the computation complexity for AVS to H.264/AVC intra transcoding. The evaluation results show that the proposed algorithm can save at least 50% computation with about 0.1 dB loss in PSNR.

**MA1-L5     Multimedia Interfaces I (Lecture)**

Time: Monday, July 10, 09:30 - 10:50

Place: Toronto II

Chair: Jie Yang, Carnegie Mellon University

09:30 - 09:50

**MA1-L5.1     SYNTHESIS AND CONTROL OF HIGH RESOLUTION FACIAL EXPRESSIONS FOR VISUAL INTERACTIONS***Chan-Su Lee, Ahmed Elgammal, Dimitris Metaxas, Rutgers University, United States*

The synthesis of facial expression with control of intensity and personal styles is important in intelligent and affective human-computer interaction, especially in face-to-face interaction between human and intelligent agent. We present a facial expression animation system that facilitates control of expressiveness and style. We learn a decomposable generative model for the nonlinear deformation of facial expressions by analyzing the mapping space between low dimensional embedded representation and high resolution tracking data. Bilinear analysis of the mapping space provides a compact representation of the nonlinear generative model for facial expressions. The decomposition allows synthesis of new facial expressions by control of geometry and expression style. The generative model provides control of expressiveness preserving nonlinear deformation in the expressions with simple parameters and allows synthesis of stylized facial geometry. In addition, we can directly extract the MPEG-4 Facial Animation Parameters (FAPs) from the synthesized data, which allows using any animation engine that supports FAPs to animate new synthesized expressions.

09:50 - 10:10

**MA1-L5.2     TOWARDS ROBUST INTUITIVE VISION-BASED USER INTERFACES***Oliver Schreer, Peter Eisert, Peter Kauff, Ralf Tanger, Fraunhofer – Heinrich Hertz Institute, Germany; Roman Englert, Deutsche Telekom Laboratories, Germany*

In future videocommunication services, the user's communication device, such as PC, laptop, PDA or mobile phone is equipped with new interaction modalities. These can be cameras and microphones on the capturing side and speech synthesis and video/3D graphics on the rendering side. Haptic and tactile interfaces become also available. These modalities help the user to interact more intuitive with complex devices and tools and provide new services. Hence, a key challenge of new modalities is robustness and stability under general conditions in arbitrary environments. Furthermore, inexperienced users should be able to use these new capabilities without dedicated knowledge of device settings or algorithms. In this paper, we will present some key components for a robust vision-based user interface, which are integrated in an advanced future videocommunication service.

10:10 - 10:30

**MA1-L5.3     A MULTIMEDIA SYSTEM FOR ROUTE SHARING AND VIDEO-BASED NAVIGATION***Wen Wu, Jie Yang, Carnegie Mellon University, United States; Jing Zhang, General Motors R&D, United States*

Trip planning and in-vehicle navigation are crucial tasks for easier and safer driving. The existing navigation systems are based on machine intelligence without allowing human knowledge incorporation. These systems give turn guidance with abstract visual instruction and have not reached the potential of minimizing driver's cognitive load, which is the amount of mental processing power required. In this paper, we describe the development of a multimedia system that makes driving and navigation safer and easier by offering tools for route sharing in trip planning and video-based route guidance during driving. The system provides a multimodal interface for a user to share his/her route with others by drawing on a digital map, naturally incorporating human knowledge into the trip planning process. The system gives driving instructions by overlaying navigational arrows onto live video and providing synthesized voice to reduce the driver's cognitive load, in addition to presenting landmark images for key maneuvers. We describe our observations which had motivated the development of the system, detailed architecture and user interfaces, and finally discuss our initial test findings in the real-road driving context.

10:30 - 10:50

**MA1-L5.4 DESIGN AND IMPLEMENTATION OF A MULTIMEDIA PERSONALIZED SERVICE OVER LARGE SCALE NETWORKS**

*Xiaorong Li, Terence Hung Gih Guang, Institute of High Performance Computing, Singapore; Bharadwaj Veeravalli, National University of Singapore, Singapore*

In this paper, we proposed to setup a distributed multimedia system which aggregates the capacity of multiple servers to provide customized multimedia services in a cost-effective way. This system enables clients to customize their services by specifying the service delay or the viewing times. We developed an experimental prototype where media servers can cooperate in streams caching, replication and distribution. We applied a variety of stream distribution algorithms to the system and studied their performance under the real-life situations with limited network resources and varying request arrival pattern. The results show such a system can provide cost-effective services and be applied to practical environments.

**MA1-L6: Monday, July 10, 09:30 - 10:50**

**MA1-L6 Multimedia Database Applications I (Lecture)**

Time: Monday, July 10, 09:30 - 10:50

Place: Varley

Chair: Alberto Del Bimbo, Università di Firenze

09:30 - 09:50

**MA1-L6.1 FINDING A SEMANTIC STRUCTURE INTERACTIVELY IN IMAGE DATABASES**

*Manjeet Rege, Ming Dong, Farshad Fotouhi, Wayne State University, United States*

We present a new approach to organize an image database by finding a semantic structure interactively based on multi-user relevance feedback. By treating user relevance feedbacks as weak classifiers and combining them together, we are able to capture the categories in the users' mind and build a semantic structure in the image database. Experiments performed on an image database consisting of general purpose images demonstrate that our system outperforms some of the other conventional methods.

09:50 - 10:10

**MA1-L6.2 3D FACE IDENTIFICATION BASED ON ARRANGEMENT OF SALIENT WRINKLES**

*Gianni Antini, Stefano Berretti, Alberto Del Bimbo, Pietro Pala, University of Firenze, Italy*

In this paper, we propose an original framework for three dimensional face representation and matching for identification purposes. Basic traits of a face are encoded by extracting curves of salient ridges and ravines from the surface of a dense mesh. A compact graph representation is then extracted from these curves through an original modeling technique capable to quantitatively measure spatial relationships between curves in a three dimensional space. In this way, face recognition is obtained by matching 3D graph representations of faces. Experimental results on a 3D face database show that the proposed solution attains high recognition accuracy and is quite robust to facial expression and pose changes.

10:10 - 10:30

**MA1-L6.3 CONTENT-FREE IMAGE RETRIEVAL USING BAYESIAN PRODUCT RULE**

*David Liu, Tsuhan Chen, Carnegie Mellon University, United States*

Content-free image retrieval uses accumulated user feedback records to retrieve images without analyzing image pixels. We present a Bayesian-based algorithm to analyze user feedback and show that it outperforms an existing maximum entropy algorithm, according to extensive experiments on trademark logo and 3D model datasets. The proposed algorithm also has the advantage of being applicable to both content-free and traditional content-based image retrieval, thus providing a common framework for these two paradigms.

**MA1-L6.4 GEODEC: ENABLING GEOSPATIAL DECISION MAKING**

*Cyrus Shahabi, Yao-Yi Chiang, Kelvin Chung, Kai-Chen Huang, Jeff Khoshgozaran-Haghighi, Craig Knoblock, Sung Chun Lee, Ulrich Neumann, Ram Nevatia, Arjun Rihan, Snehal Thakkar, Suya You, University of Southern California, United States*

The rapid increase in the availability of geospatial data has motivated the effort to seamlessly integrate this information into an information-rich and realistic 3D environment. However, heterogeneous data sources with varying degrees of consistency and accuracy pose a challenge to such efforts. We describe the Geospatial Decision Making (GeoDec) system, which accurately integrates satellite imagery, three-dimensional models, textures and video streams, road data, maps, point data and temporal data. The system also includes a glove-based user interface.

### MA1-P1: Monday, July 10, 09:30 - 10:30

**MA1-P1 Media Coding (Poster)**

Time: Monday, July 10, 09:30 - 10:30

Place: Toronto III

Chair: Lap-Pui Chau, Nanyang Technological University

**MA1-P1.1 MULTI-VIEW VIDEO CODING USING VIEW INTERPOLATION AND REFERENCE PICTURE SELECTION**

*Masaki Kitahara, NTT Corporation, Japan; Hideaki Kimata, NTT Advanced Technology Corporation, Japan; Shinya Shimizu, Kazuto Kamikura, Yoshiyuki Yashima, NTT Corporation, Japan; Kenji Yamamoto, Tomohiro Yendo, Toshiaki Fujii, Masayuki Tanimoto, Nagoya University, Japan*

We propose a new multi-view video coding method using adaptive selection of motion/disparity compensation based on H.264/AVC. One of the key points of the proposed method is the use of view interpolation as a tool for disparity compensation by assigning reference picture indices to interpolated images. Experimental results have shown that significant gains can be obtained compared to the conventional approach that was often used.

**MA1-P1.2 A NEW ORIENTED ADAPTIVE CROSS SEARCH ALGORITHM FOR BLOCK MATCHING MOTION ESTIMATION**

*Heng Yang, Qing Wang, Northwestern Polytechnical University, China*

Block-matching motion estimation plays an important role in video coding and faster, more robust and more effective search algorithms are needed. Recently, a great number of fast block matching algorithms (BMAs) have been proposed in the literature based on the discovery of the center-biased characteristics of motion-vector distribution. In this paper, a novel oriented adaptive cross search (OACS) algorithm is proposed, where small cross, large cross and T-shape search patterns are defined and utilized adaptively. In accordance with the adaptive tracing of orientation change or optimal point, three kinds of key points are defined to decide which kind or which oriented pattern may be chosen for the next step. Experimental results on the benchmarks have shown that the OACS algorithm can provide average speed ups of 74.65%, 39.78%, 42.44%, and 7.84% over DS, SDS, CDS, and SCDS, respectively. Finally, the mean absolute distortion and PSNR of luminance component are close to the results of other fast BMAs and the similar search accuracy can be maintained as expected.

### **MA1-P1.3 OPTIMIZED RATE ALLOCATION FOR UNBALANCED MULTIPLE DESCRIPTION VIDEO CODING OVER UNRELIABLE PACKET NETWORK**

*Bin Li, Feng Huang, Lifeng Sun, Yang Shiqiang, Tsinghua University, China*

Video transmission over unreliable packet network is in general hampered by the packet losses and constraint by stringent playback deadline. With these two key factors in consideration, Multiple Description Coding (MDC), comprising balanced and Unbalanced MDC has been proposed as an error-robust source coding technique. Recently, transmitting multiple descriptions over a single path is interesting due to the unavailability of multiple independent paths. Therefore, in this paper, we investigate the problem of rate allocation for the High-Resolution (HR) and Low-Resolution (LR) descriptions in UMDC transmission over single path. We first propose an approximate but efficient rate allocation model with the aid of two-state Markov link model and a simple distortion model at the sender side. Then we conduct extensive experiments to verify the proposed model and more excitedly the simulation results clearly demonstrate the effectiveness of proposed model.

### **MA1-P1.4 VIDEO TRANSCODING FOR PACKET LOSS RESILIENCE BASED ON THE MULTIPLE DESCRIPTIONS**

*Il Koo Kim, Nam Ik Cho, Seoul National University, Republic of Korea*

This paper proposes error resilient video transcoding structures based on the multiple description(MD) scheme. Two structures are proposed for different use, namely low complexity simple MD transcoding structure(Simple MD) and adaptive MD transcoding structure(Adaptive MD). Simple MD structure always converts and splits the given bit stream into two descriptions regardless of channel condition. On the other hands, the Adaptive MD structure considers the channel condition, and it switches between the MD and single description(SD) mode depending on the packet loss rate(PLR). In the switching process, the best mode which produces less end-to-end cost is chosen between SD and MD under given channel condition. The expected end-to-end cost is optimally estimated based on the rate-distortion theory at the time of encoding. The simulation results show that the Simple MD structure provides efficient performance than the conventional error resilient transcoding structures based on the spatial and temporal error localization. And the Adaptive MD structure shows higher PSNR than the conventional methods under the comparable complexity.

### **MA1-P1.5 A NEW QUANTIZATION FOR RATE CONTROL WITH FRAME VARIATION CONSIDERATION**

*Seonki Kim, LG Electronics, Inc., Republic of Korea; Seong-Jae Lim, Gwangju Institute of Science and Technology, Republic of Korea*

The main role of rate control in video coding is to handle a trade-off between a bandwidth limit and a video quality. In the low-delay video communication systems, its responsibility is more important. In this work, we propose a new rate control algorithm. In this paper, we have two approaches: one is a consideration of variations among the frames at the framelevel bit allocation, and the other is a model-based adaptive quantization at the macroblock-level. The proposed scheme is done with a low computational complexity. For simulations, we implement the proposed algorithm to H.263 video codec since it is widely used as a video coding tool in real-time mobile AV systems. The proposed scheme shows more improved picture quality than H.263 TMN8 and generates coding bits close to the target bits. The results also show more regular PSNR fluctuation than H.263 TMN8.

### **MA1-P1.6 MOTION ALIGNED SPATIAL SCALABLE VIDEO CODING**

*Debing Liu, Chinese Academy of Sciences, China; Yuwen He, Shipeng Li, Microsoft Research Asia, China; Debin Zhao, Wen Gao, Chinese Academy of Sciences, China*

A motion aligned spatial scalable video coding scheme (MA-SSC) is proposed in this paper. Different from the traditional spatial scalable coding schemes derived from MPEG-2, in the proposed scheme only one set of intra or inter prediction modes are optimally selected by jointly considering the base and enhancement layers. Thus, it saves one set of macroblock (MB) mode and motion vectors. Moreover, the combined motion estimation can reduce the residual coding bits of the base layer. The MA-SSC and traditional spatial scalable coding schemes are both implemented based on H.264 reference software to evaluate their performance. Simulation results show that the enhancement layer coding efficiency of MA-SSC is up to 0.6dB better than that of the traditional scheme, while the base layer coding efficiency of MA-SSC decreases less than 0.3db compared with the single-layer coding.

#### **MA1-P1.7 FAST MULTI-REFERENCE FRAME MOTION ESTIMATION VIA DOWNHILL SIMPLEX SEARCH**

*Chen-Kuo Chiang, Shang-Hong Lai, National Tsing Hua University, Taiwan*

Multi-reference frame motion estimation improves the accuracy of motion compensation in video compression, but it also dramatically increases computational complexity. Based on tracing motion vector trajectories, fast approximated motion estimation results can be obtained for multi-reference frames. In this paper, we extend the downhill simplex search to multiple reference frames and propose several enhanced schemes to improve its efficiency and accuracy. Experimental results show that the proposed algorithm outperforms several representative single-reference frame block matching methods.

#### **MA1-P1.8 FAST PROGRESSIVE MODEL REFINEMENT GLOBAL MOTION ESTIMATION ALGORITHM WITH PREDICTION**

*Haifeng Wang, Jia Wang, Qingshan Liu, Hanqing Lu, Chinese Academy of Sciences, China*

Global Motion Estimation (GME) is an important part in the object-based applications. In this paper, a Fast Progressive Model Refinement (FPMR) GME algorithm is proposed. It can select the appropriate motion model according to the complexity of the camera motion. Two techniques are used to accelerate the procedure of FPMR. The first is an outlier prediction based feature point selection method. It can predict outliers from that of the last frame and therefore can effectively remove the influence of outliers on parameter calculation. The second is an intermediate-level model prediction method, which is used to fast the model selection and the parameter calculation procedure. Experiments show that the proposed algorithm is above two times faster than that of the Feature-based Fast and Robust GME technique.

#### **MA1-P1.9 LOW-COMPLEXITY ADAPTIVE BLOCK-SIZE TRANSFORM BASED ON EXTENDED TRANSFORMS**

*Honggang Qi, Wen Gao, Chinese Academy of Sciences, China; Siwei Ma, University of Southern California, United States; Debin Zhao, Xiangyang Ji, Chinese Academy of Sciences, China*

In this paper, a low-complexity 8x8/4x4 Adaptive Block-size Transform (ABT) scheme is tentatively proposed for the Chinese Audio and Video coding Standard (AVS). In the proposed ABT scheme, an integer 8x8 transform is derived from the integer 4x4 transform used in AVS coding standard according to an transform extension principle. It not only has high energy compacted property but also can be implemented with several additions and shifts, and all intermediate results are limited with 16-bit. The 8x8 transform and 4x4 transform can be merged together and share the same scale matrix so that the hardware units and storage resources are efficiently saved for both encoder and decoder. The experimental results on numerous sequences show that the proposed ABT scheme can achieves significant performance improvement for AVS.

### **MA1-P2: Monday, July 10, 09:30 - 10:30**

#### **MA1-P2 Media Coding and Processing II (Poster)**

Time: Monday, July 10, 09:30 - 10:30

Place: Toronto III

Chair: Hiroshi Watanabe, Waseda University

#### **MA1-P2.11 MODIFIED WINNER UPDATE WITH ADAPTIVE BLOCK PARTITION FOR FAST MOTION ESTIMATION**

*Shou-Der Wei, Shao-Wei Liu, Shang-Hong Lai, National Tsing Hua University, Taiwan*

Motion estimation (ME) plays an important role in video compression. Block-based ME has been adopted in most video compression standards due to its efficiency. In this paper, we propose a novel and fast block-based ME algorithm based on applying the modified winner-update scheme in conjunction with the adaptive partition order of macroblock. The partition order is determined from the block gradient distribution. Experimental results show the proposed algorithm achieves the optimal motion estimation very efficiently.

## **MA1-P2.12 FAST ADAPTATION DECISION TAKING FOR CROSS-MODAL MULTIMEDIA CONTENT ADAPTATION**

*Martin Prangl, Hermann Hellwagner, Klagenfurt University, Austria; Tibor Szkaliczki, Hungarian Academy of Sciences, Hungary*

In order to enable transparent and convenient use of multimedia content across a wide range of networks and devices, content adaptation is an important issue within multimedia frameworks. The so called Digital Item Adaptation (DIA) standard is one of the core concepts of the MPEG-21 framework that will support the adaptation of multimedia resources according to device capabilities, underlying network characteristics, and user preferences. Most multimedia adaptation engines for providing Universal Multimedia Access (UMA) scale the content with respect to terminal capabilities and resource constraints. This paper focuses on the cross-modal adaptation decision taking process considering the user environment and terminal capabilities as well as resource limitations on the server, network, and client side. This approach represents a step toward increased Universal Multimedia Experience (UME). Based on four different algorithms for solving this optimization process, we present an evaluation of results gained by running their implementations on different test networks.

## **MA1-P2.13 AN EFFICIENT MEMORY CONSTRUCTION SCHEME FOR AN ARBITRARY SIDE GROWING HUFFMAN TABLE**

*Sung-Wen Wang, Shang-Chih Chuang, Chih-Chieh Hsiao, Yi-Shin Tung, Ja-Ling Wu, National Taiwan University, Taiwan*

By grouping the common prefix of a Huffman tree, in stead of the commonly used single-side rowing Huffman tree (SGH-tree), we construct a memory efficient Huffman table on the basis of an arbitrary-side growing Huffman tree (AGH-tree) to speed up the Huffman decoding. Simulation results show that, in Huffman decoding, an AGH-tree based Huffman table is 2.35 times faster than that of the Hashemian's method (an SGH-tree based one) and needs only one-fifth the corresponding memory size. In summary, a novel Huffman table construction scheme is proposed in this paper which provides better performance than existing construction schemes in both decoding speed and memory usage.

## **MA1-P2.14 A STUDY ON SPATIAL SCALABLE CODING USING VECTOR REPRESENTATION**

*Yuki Yamamoto, Kei Kawamura, Hiroshi Watanabe, Waseda University, Japan*

The major advantage of vector representation of an image is that the image quality is maintained for arbitrary scaling. In recent years, a demand for scalable image coding has been increasing because of the wide variety of available digital contents and display terminals. Conventional scalable coding schemes are based on raster representation, and thus, line drawings deteriorate in quality when expanded and shrunk. In this paper, we propose an edge reconstruction method using vector representation for the purpose of keeping a consistent spatial scalability on transmission and display. We take an anti-aliasing into account in edge areas for approximation of luminance values around the edge. The proposed method can improve PSNR by up to 2 dB as compared to the conventional methods when image is expanded and shrunk.

## **MA1-P2.15 NON-BANDLIMITED RESAMPLING OF IMAGES**

*Beilei Huang, Nanyang Technological University, Singapore; Edmund M-K Lai, Massey University, New Zealand*

The resampling of discrete-time signals where the underlying analog signal is non-bandlimited is considered in this paper. We extend the generalized sampling theory developed based on the principle of consistency to resampling. Realizing the resampling system has both discrete input and output, the performance of the resampling filter is considered in  $\mathcal{L}^2$  instead of the traditionally used  $\mathcal{L}^\infty$ . We show that the performance of the resampling system depends on the resampling rate instead of the actual interpolating kernels. The theory can be applied to image processing applications like zooming to provide better response to high frequency components. Since the resampling process is discrete in nature, our filter designed to optimize resampling in  $\mathcal{L}^2$  is shown to outperform other techniques designed in  $\mathcal{L}^\infty$ .

## **MA1-P2.16 EFFICIENT MOTION ESTIMATION USING SORTING-BASED PARTIAL DISTORTION SEARCH**

*Chou-Chen Wang, Chia-Jung Lo, Cheng-Wei Yu, I-Shou University, Taiwan*

An efficient motion-estimation algorithm based on partial block distortion using sorted significant features including bit-plane and mean is proposed. The proposed algorithm can obtain relatively accurate motion vectors with a reduced computational load. Simulation results show that the proposed method achieves its MSE performance very close to the full search method, while requiring only 6-8% of the computation needed by the full search. Furthermore, the performance of our method is better than other algorithms based on partial block distortion search.



#### **MA1-P2.17 SEMANTIC 3D FACE MESH SIMPLIFICATION FOR TRANSMISSION AND VISUALIZATION**

*Alessandro Colombo, Claudio Cusano, Raimondo Schettini, University of Milano, Bicocca, Italy*

Three-dimensional data generated from range scanners is usually composed of a huge amount of information. Simplification and compression techniques must be adopted in order to reduce transmission or processing time and to allow real-time visualization. In this paper we propose an approach to semantic simplification of triangular meshes representing faces. The algorithm is aimed to preserve facial features and can be used especially for face recognition systems purposes. In a first phase we detect salient regions using a 3D face detector based on curvature analysis and holistic classification. In a second phase, vertex decimation is applied to the mesh with different decimation parameters for salient and non salient-regions. We have tested our algorithm on a set of 150 acquisitions obtaining good visual quality meshes with approximately 90% or more of decimated vertexes.

#### **MA1-P2.18 AUTOMATIC EXTRACTION OF GEOMETRIC LIP FEATURES WITH APPLICATION TO MULTI-MODAL SPEAKER IDENTIFICATION**

*Ivana Arsic, Roger Vilagut, Jean-Philippe Thiran, Swiss Federal Institute of Technology, Lausanne, Switzerland*

In this paper we consider the problem of automatic extraction of the geometric lip features for the purposes of multi-modal speaker identification. The use of visual information from the mouth region can be of great importance for improving the speaker identification system performance in noisy conditions. We propose a novel method for automated lip features extraction that utilizes color space transformation and a fuzzy-based c-means clustering technique. Using the obtained visual cues closed-set audio-visual speaker identification experiments are performed on the CUAVE database, showing promising results.

#### **MA1-P2.19 AN OPTIMAL NON-UNIFORM SCALAR QUANTIZER FOR DISTRIBUTED VIDEO CODING**

*Bo Wu, Chinese Academy of Sciences, China; Xun Guo, Debin Zhao, Wen Gao, Harbin Institute of Technology, China; Feng Wu, Microsoft Research Asia, China*

In this paper, we propose a novel algorithm to design an optimal non-uniform scalar quantizer for distributed video coding, which aims at achieving a coding rate close to joint conditional entropy of the quantized video frames given the side information. Wyner-Ziv theory on source coding is employed as the basic coding principle and the asymmetric scenario is considered. In this algorithm, a probability distribution model, which considers the influence of the joint distribution of input source and side information to the coding performance, is established and used as the optimality condition firstly. Then, a modified Lloyd Max algorithm is used to design the scalar quantizer to give an optimal quantization for input source before coding. Experimental results show that proposed algorithm is efficient for Gaussian source and practical video coding. Compared to uniform scalar quantization, proposed algorithm can improve coding performance largely, especially at low bit rate.

#### **MA1-P2.20 ACOUSTIC ECHO CANCELLATION IN A CHANNEL WITH RAPIDLY VARYING GAIN**

*Sumit Basu, Microsoft Research, United States*

We present a method for performing acoustic echo cancellation in a channel with rapidly varying gain and thus a rapidly varying channel characteristic. This is a situation in which standard AEC approaches perform poorly. Our method involves learning a scale-free channel characteristic ( $H_{\sim}$ ). We then apply this to a windowed version of the signal and remove the projection of the transformed signal from the output signal. We also develop a “ramp projection” method that allows for a linear variation in gain within the window. We show results in a telephony application with 3 dB to more than 8 dB of improvement over conventional AEC using the simple projection and an additional 1 dB using the ramp projection.



**MA1-P3 Stereo, Multiview and Media Analysis (Poster)**

Time: Monday, July 10, 09:30 - 10:30

Place: Toronto III

Chair: Nicolas Georganas, University of Ottawa

**MA1-P3.21 METHODS FOR NONE INTRUSIVE DELAY MEASUREMENT FOR AUDIO COMMUNICATION OVER PACKET NETWORKS**

*Mohammad Zad-Issa, Norbert Rossello, Laurent Pilati, Mindspeed Technologies, United States*

Measurement of the delay is an important and common problem in communication over packet networks. The end-to-end and the round trip delay are among the factors directly impacting the quality of service as well as the user satisfaction. Multimedia gateways or base stations that perform echo cancellation or suppression often rely on the round trip delay to enhance their performance or to reduce the computational complexity of echo processing logics. In this work, we present two none intrusive methods for delay estimation and tracking. Both methods find the delay using the actual audio signal that is sent through the network. The first approach uses the MDCT transformed domain coefficients of the signal while the second operates in a perceptual domain. Experiments illustrate that both schemes can track the end-to-end and the round trip delay under various network and signal conditions.

**MA1-P3.22 ADAPTIVE DUAL AK-D TREE SEARCH ALGORITHM FOR ICP REGISTRATION APPLICATIONS**

*Jiann-Der Lee, Shih-Sen Hsieh, Chung-Hsien Huang, Li-Chang Liu, Chang Gung University, Taiwan; Chien-Tsai Wu, Shin-Tseng Lee, Jyi-Feng Chen, Chang Gung Memorial Hospital, Taiwan*

An algorithm for finding coupling points plays an important role in the Iterative Closest Point algorithm (ICP). In recent researches of finding coupling points, An algorithm for finding coupling points plays an important role in the Iterative Closest Point algorithm (ICP) which is widely used in registration applications in medical and 3-D architecture areas. In recent researches of finding coupling points, Approximate K-D tree search algorithm (AK-D tree) is an efficient nearest neighbor search algorithm with comparable results. We proposed Adaptive Dual AK-D tree search algorithm (ADAK-D tree) for searching and synthesizing coupling points as significant control points to improve the registration accuracy in ICP registration applications. ADAK-D tree utilizes AK-D tree twice in different geometrical projection orders to reserve true nearest neighbor points used in later ICP stages. An adaptive threshold in ADAK-D tree is used to reserve sufficient coupling points for a smaller alignment error. Experimental results are shown that the registration accuracy of using ADAK-D tree is improved more than the result of using AK-D tree and the computation time is acceptable.

**MA1-P3.23 APPROXIMATED CORRELATION MATRIX AND PULSE PREDICTION FOR FAST ALGEBRAIC CODE-EXCITED LPC SPEECH CODERS**

*Shu-Min Tsai, Jar-Ferr Yang, National Cheng Kung University, Taiwan*

An efficient algebraic-code-excited linear prediction (ACELP) search mechanism based on an approximated correlation matrix of the vocal impulse response is proposed to reduce the computation. By exploring statistical characteristics, we only need to calculate a small portion of correlation coefficients before ACELP search procedure. If we further combine a pulse position prediction algorithm, we can reduce the arithmetic complexity in pre-computing autocorrelation matrix and the number of pulse position combinations. Simulations show that the proposed method can effectively reduce the computation in ACELP codebook search while the performance degradation is imperceptible.

**MA1-P3.24 SOUND SOURCE SEPARATION OF TRIO USING STEREO MUSIC SOUND SIGNAL BASED ON INDEPENDENT COMPONENT ANALYSIS**

*Satoru Morita, Yasuhito Nanri, Yamaguchi University, Japan*

It is necessary that the number of the observation signals equals to the number of source signals, if independent component analysis is used to perform the sound source separation. It is difficult to perform sound source separation from a stereo music sound signal when the number of sound sources are more than two. We propose the technique to perform sound source separation from a stereo music sound signal that the number of sound sources is more than two using the frequency analysis and independent component analysis.

### **MA1-P3.25 INTERACTIVE MULTI-VIEW VISUAL CONTENTS AUTHORING SYSTEM**

*Injae Lee, Myungseok Ki, Seyoon Jeong, Electronics and Telecommunications Research Institute (ETRI), Republic of Korea; Kyuheon Kim, Kyunghee University, Republic of Korea*

This paper describes issues and consideration on authoring of interactive multi-view visual content based on MPEG-4. The issues include types of multi-view visual content; functionalities for user-interaction; scene composition for rendering; and multi-view visual content file format. The MPEG-4 standard, which aims to provide an object based audiovisual coding tool, has been developed to address the emerging needs from communications, interactive broadcasting as well as from mixed service models resulting from technological convergence. Due to the feature of object based coding, the use of MPEG-4 can resolve the format diversity problem of multi-view visual contents while providing high interactivity to users. Throughout this paper, we will present which issues need to be determined and how currently available tools can be effectively utilized for interactive multi-view visual content creation.

### **MA1-P3.26 DYNAMIC VIEW-DEPENDENT MULTIREOLUTION TERRAIN VISUALIZATION**

*Din-Chang Tseng, Chung-Chieh Huang, National Central University, Taiwan*

Large-area terrain visualization is important to multimedia and military applications. However, such a huge amount of terrain data is not easily able to process on a general personal computer. Thus, how to efficiently process and display a large-area terrain data is the main challenge. In this study, we combine the proposed techniques of multiresolution terrain modeling, view-dependent rendering, and dynamic loading with the digital terrain models and satellite images to dynamically browse the large terrain models and scene environment on a general personal computer.

### **MA1-P3.27 3D MODEL SEARCH BASED ON STOCHASTIC ARG MATCHING**

*Naoto Nakamura, Yoshihiro Okada, Koichi Nijima, Kyushu University, Japan*

Due to the high performance of recent computer graphics hardware, 3D CG and CG animations have become in great demand for various applications. Many 3D models have already been created and stored. We need any 3D model search system that allows us to retrieve our required 3D models accurately. In this paper, the authors propose a 3D model search system that uses the attributed relational graph (ARG) of each 3D model, and that employs the stochastic ARG matching to measure the similarity among them. The proposed system gives better search results than those of D2 method in their experiments. This paper also describes what kinds of features of a 3D model are used as each vertex attributes for the stochastic ARG matching.

### **MA1-P3.28 OPTIMAL ERROR PROTECTION OF PROGRESSIVELY COMPRESSED 3D MESHES**

*Shakeel Ahmad, Raouf Hamzaoui, University of Konstanz, Germany*

Given a number of available layers of source data and a transmission bit budget, we propose an algorithm that determines how many layers should be sent and how many protection bits should be allocated to each transmitted layer such that the expected distortion at the receiver is minimum. The algorithm is used for robust transmission of progressively compressed 3D models over a packet erasure channel. In contrast to the previous approach, which uses exhaustive search, the time complexity of our algorithm is linear in the transmission bit budget.

### **MA1-P3.29 A NEW IMAGE CORRECTION METHOD FOR MULTIVIEW VIDEO SYSTEM**

*Feng Shao, Gangyi Jiang, Mei Yu, Ningbo University, China; Xiexiong Chen, Zhejiang University, China*

Because of scene illumination or camera calibration, color appearance of the same object between different viewpoints may be different in multiview video system. Traditional illumination compensation algorithm for image is unable to solve this problem effectively. In this paper, a novel color correction method for multiview video system is proposed based on retinex color constancy theory. To eliminate influence of un-consistent light sources, histogram equalization, retinex processing and color restoration are performed for multiview images to extract reflectance that describes object intrinsic properties. Experimental results show that the proposed image correction method for multiview video system is effective.

### **MA1-P3.30 LOCALLY REFINED COLLISION DETECTION OF LARGE SCALE COMPLEX POLYGONAL MESHES IN DISTRIBUTED VIRTUAL ENVIRONMENTS**

*Peiran Liu, Nicolas Georganas, University of Ottawa, Canada; Gerhard Roth, National Research Council of Canada, Canada*

This paper presents a new locally refined collision detection approach for large scale complex meshes in distributed virtual environments (DVEs) where exact and interactive interference detection is required. Transmitting models with millions of polygons is time consuming in comparison with transmitting simple models. Even if the models are transmitted in progressive manner, the earlier received models are in low level-of-detail (LOD) because the models are refined globally. Increasing the accuracy of collision detection (CD) at the client still takes time because increasing the LOD of the model is a slow process. The new approach is composed of an AB-tree collision query algorithm and a new mesh refinement algorithm on a space partitioned mesh (SPM) representation. It deals with this problem by selectively refining the models at certain areas that are predicted to collide with other objects and transmitting the refined parts instead of the entire model from server to client. The accuracy of CD is increased quickly at the predicted contact areas. An interactive rate is guaranteed by reducing network response time and dynamically adjusting the complexity and the space cost of the collision query algorithm.

### **MA1-P3.31 SCALABLE MULTI-VIEW VIDEO CODING FOR INTERACTIVE 3DTV**

*Nukhet Ozbek, Ege University International Computer Institute, Turkey; A. Murat Tekalp, Koç University, Turkey*

A standard for scalable video coding (SVC) is currently being worked on by the ISO MPEG Group. Work on standardization of multiple-view video coding (MVC) has also recently started under the ISO MPEG. Although there are many approaches published on SVC and MVC, there is no current work reported on scalable multi-view video coding (SMVC). This paper presents new coding structures for scalable stereo and multi-view video coding. The proposed structures are implemented as extensions to the JSVM software and resulting bitrates and PSNR are demonstrated. SMVC can be used for transport of multiview video over IP for interactive 3DTV by dynamic adaptive combination of temporal, spatial, and SNR scalability according to network conditions.

### **MA1-P3.32 EXTRACTION OF OUTCROP POINTS FROM VISUAL HULLS FOR MOTION ESTIMATION**

*Masahiro Toyoura, Masaaki Iiyama, Koh Kakusho, Michihiko Minoh, Kyoto University, Japan*

In this article, we discuss 3D shape reconstruction of an object in a rigid motion with the volume intersection method. When the object moves rigidly, the cameras change their relative positions to the object at every moment. To estimate the motion correctly, we propose new feature points called outcrop points on the reconstructed 3D shape. These points are guaranteed to be located on the real surface of the object. If the rigid motion of the object can be correctly estimated, cameras at different moments serve as the cameras in different positions virtually. With these cameras in time sequences, we can increase accuracy of the reconstructed 3D shape without increasing the number of cameras. Based on this idea, we reconstruct an accurate shape of the object in motion from images obtained by the certain number of cameras. As the result, we can acquire an accurate shape from images in time sequences.

## **MA1-P4: Monday, July 10, 09:30 - 10:30**

### **MA1-P4 Media Coding and Processing I (Poster)**

Time: Monday, July 10, 09:30 - 10:30

Place: Toronto III

Chair: Joern Ostermann, University of Hannover

### **MA1-P4.33 RELATIVE DEPTH LAYER EXTRACTION FOR MONOSCOPIC VIDEO BY USE OF MULTIDIMENSIONAL FILTER**

*Jing-Ying Chang, Chao-Chung Cheng, Shao-Yi Chien, Liang-Gee Chen, National Taiwan University, Taiwan*

This paper presents a relative depth layer extraction system for monoscopic video, using multi-line filters and a layer selection algorithm. Main ideas are to extract multiple linear trajectory signals from videos and to determine their relative depths using the concept of motion parallax. The proposed superficial line model used for detecting slow moving objects provides sufficient taps within few frames to reduce frame buffer, while the closest-hit line model used for detecting fast motion objects provides few enough taps to prevent blurring. To increase the correctness of layer map, three-level layer map co-decision is used to compensate low texture region defect.

#### **MA1-P4.34 ROBUST RIGID HEAD MOTION ESTIMATION BASED ON DIFFERENTIAL EVOLUTION**

*Axel Weissenfeld, Onay Urfalioglu, Kang Liu, Joern Ostermann, University of Hannover, Germany*

In this paper we present a system to robustly estimate the 3D position of a human head. Before the face model is positioned in the initial frame, it is adapted to the 3D scan of the tracked human head. Head tracking is achieved by minimizing a robust cost function with a stochastic optimization algorithm called Differential Evolution. This approach enables the estimation of large motions between consecutive frames. Furthermore, the algorithm can even handle a large number of outliers e.g. caused by occlusion and still estimate the precise position.

#### **MA1-P4.35 REAL-TIME ITERATIVE SPECTRUM INVERSION WITH LOOK-AHEAD**

*Xinglei Zhu, Institute for Infocomm Research, Singapore; Gerald Beauregard, muvee Technologies, Singapore; Lonce Wyse, Institute for Infocomm Research, Singapore*

In this paper, we present an algorithm for Real-time Iterative Spectrogram Inversion (RTISI) with Look-Ahead (RTISI-LA). RTISI-LA reconstructs a time-domain signal from a given sequence of short-time Fourier transform magnitude (STFTM) spectra without phase information. Whereas RTISI reconstructs the current frame using only magnitude spectra information for previous frames and the current frame, RTISI-LA also uses magnitude spectra for a small number future frames. This allows RTISI-LA to achieve substantially higher signal-to-noise (SNR) performance than either RTISI or the traditional iterative method proposed by Griffin & Lim with an equivalent computational load, while retaining the real-time properties of RTISI.

#### **MA1-P4.36 AN OPTIMAL AND STATISTICALLY ROBUST CORRELATION TECHNIQUE FOR BLOCK BASED MOTION ESTIMATION**

*Fedwa Essannouni, GSCM, Morocco; Rachid Oulad Haj Thami, Laboratoire SI2M, Equipe WiM, Morocco; Ahmed Salam, LMPA, France; Driss Aboutajdine, GSCM, Morocco*

In this paper we propose an optimal and robust correlation technique for the local motion estimation purposes. It is based on the maximization of a statistical robust matching function, which is computed in the frequency domain and therefore can be implemented by fast transformation algorithms. We show that our method achieves a significant speed up and robustness over the full search block-matching algorithm. We also present a comparative performance analysis, which shows that the proposed method greatly outperforms the state-of-the-art in correlation motion estimation.

#### **MA1-P4.37 HUMAN VISION SYSTEM AWARE EXHAUSTIVE BLOCK-MATCHING ALGORITHM**

*David Chih-Che Lin, Paul M. Chau, University of California, San Diego, United States*

In this work, homomorphic image modeling is used to make the exhaustive block-matching algorithm (EBMA) more human vision system (HVS) aware, thus yielding visually pleasing sequences. Homomorphic signal processing is used to separate the luminous and the structural components of each frame; EBMA is then applied to both components to capture the luminous and the structural changes. The combination of these two techniques is utilized to simulate the structure preserving nature of HVS. Temporal irrelevancies are reduced by removing the excess motion fields from both components. The homomorphic EBMA (H-EBMA) processed sequences have superior visual quality despite mean PSNR values that are comparable to sequences produced by traditional EBMA.

#### **MA1-P4.38 PERFORMANCE OF OPTICAL FLOW TECHNIQUES ON GRAPHICS HARDWARE**

*Marko Durkovic, Michael Zwick, Florian Obermeier, Klaus Diepold, Technische Universität München, Germany*

Since graphics cards have become programmable the recent years, numerous computationally intensive algorithms have been implemented on the now called General Purpose Graphics Processing Units (GPGPUs). While the results show that GPGPUs regularly outperform CPU based implementations, the question arose how optical flow algorithms can be ported to graphics hardware. To answer the question, the optimal algorithm structure to maximize the performance gained by using graphics cards has to be found. In this paper we compare the performance of two algorithms that are highly different in structure, implemented on both CPU and graphics hardware. Analyzing the results of the CPU and GPGPU implementation, we explore the mapping of the algorithms to the graphics hardware and thereof extract information about a preferred structure of optical flow algorithms for GPGPU based implementation.

#### **MA1-P4.39 DECENTRALIZED MULTIPLE CAMERA MULTIPLE OBJECT TRACKING**

*Wei Qu, Dan Schonfeld, University of Illinois at Chicago, United States; Magdi Mohamed, Motorola Labs, United States*

In this paper, we present a novel decentralized Bayesian framework using multiple collaborative cameras for robust and efficient multiple object tracking with significant and persistent occlusion. This approach avoids the common practice of using a complex joint state representation and a centralized processor for multiple camera tracking. When the objects are in close proximity or present multi-object occlusions in a particular camera view, camera collaboration between different views is activated in order to handle the multi-object occlusion problem. Specifically, we propose to model the camera collaboration likelihood density by using epipolar geometry with particle filter implementation. The performance of our approach has been demonstrated on both synthetic and real-world video data.

#### **MA1-P4.40 WHERE WAS THE PICTURE TAKEN: IMAGE LOCALIZATION IN ROUTE PANORAMAS USING EPIPOLAR GEOMETRY**

*Saad Khan, Fahd Rafi, Mubarak Shah, University of Central Florida, United States*

Finding the location where a picture was taken is an important problem for a variety of applications including surveying, interactive traveling and homeland security among others. The task becomes intractable though when the area under investigation reaches city/town size. The amount of data (pictures/videos) required to visually map a city, comprehensively, can be exhaustive for most search algorithms. In this paper we propose a novel method to effectively tackle this problem. The area is visually mapped as route panoramas that provide a compact yet comprehensive representation of the buildings and landmarks in the area. Given a query image taken at an arbitrary location in the area, we show that we can accurately recover the location of the camera by finding its epipole in the route panorama of the scene. To this end we show that there exists a fundamental matrix between a route panorama and a perspective image of the same scene. The fundamental matrix is calculated using feature matches as correspondences between the query image and the route panorama.

#### **MA1-P4.41 COLOR DATA CODING FOR THREE-DIMENSIONAL MESH MODELS CONSIDERING CONNECTIVITY AND GEOMETRY INFORMATION**

*Young-Suk Yoon, Electronics and Telecommunications Research Institute (ETRI), Republic of Korea; Sung-Yeol Kim, Yo-Sung Ho, Gwangju Institute of Science and Technology, Republic of Korea*

In this paper, we propose a new predictive coding scheme for color data of three-dimensional (3-D) mesh models. We exploit connectivity and geometry information to improve coding efficiency. After ordering all vertices in a 3-D mesh model with a connectivity coding technique, we propose a geometry predictor to compress the color data efficiently. The predicted color can be obtained by a weighted sum of reconstructed colors for adjacent vertices using both angles and distances between the current vertex and adjacent vertices. Simulation results show that the proposed scheme provides enhanced coding efficiency over previous works for various 3-D mesh models.

#### **MA1-P4.42 REMOVING SHADOWS USING FLASH/NOFLASH IMAGE EDGES**

*Mark S. Drew, Cheng Lu, Simon Fraser University, Canada; Graham Finlayson, University of East Anglia, United Kingdom*

Flash/noflash pairs have been used for noise-reduction in ambient-light images. But not explicitly studied is the problem of shadows in the ambient images. While shadows are lessened in a flash image, other problems arise, and other shadows are produced. It is known that we can in fact produce a flash-only (no ambient) image by subtracting the two images, but the result is not as pleasant as the ambient image, because of several artifacts due to the flash. Here, we use the pure-flash image to detect the ambient shadows. We argue that first going to a "spectrally sharpened" color space, and then focusing on the difference in a log domain of the flash image minus the ambient image, gives a very simple feature space consisting of two components -- one in an illuminant-change 3-vector direction, and one along the gray axis. This space provides excellent separation of the shadow and nonshadow areas. Inserting edges from the flash image within the ambient-shadow region into the ambient image edge map and inverting Poisson's equation fills in the shadow. In addition, by manipulating the gradient field at the shadow edges, the difference caused by shadow edges can be removed. In this way, we arrive at an image with the advantages of the ambient-only image -- warmth, no flash effects such as disturbing illumination dropoff with distance, pixel saturation etc. -- but no shadows.

**MA2-L1 Feature Extraction and Segmentation II (Lecture)**

Time: Monday, July 10, 10:50 - 12:10

Place: Carmichael

Chair: Langis Gagnon, CRIM

10:50 - 11:10

**MA2-L1.1 A TWO-LAYER GRAPHICAL MODEL FOR COMBINED VIDEO SHOT AND SCENE BOUNDARY DETECTION**

*Marc Al-Hames, Stefan Zettl, Frank Wallhoff, Stephan Reiter, Bjoern Schuller, Gerhard Rigoll, Technische Universität München, Germany*

In this work we present a novel two-layer hybrid Graphical model for combined shot and scene boundary detection in videos. In the first layer of the model, low-level features are used to detect shot boundaries. The shot layer is connected to a higher layer that detects scene or chapter boundaries from semantic features. With this structure, the model optimises the alignment for both layers at the same time and the detection results are interconnected. Experimental results on real video data show, that both layers highly benefit from this sharing of information. Compared to a baseline threshold method with the same features, the F-measure result for the shot detection has been improved by 12.6% absolute. For the scene boundary detection, the result has been improved by more than 11% absolute.

11:10 - 11:30

**MA2-L1.2 A SNAKE-BASED SEGMENTATION ALGORITHM FOR OBJECTS WITH BOUNDARY CONCAVITIES**

*Shin-Hyoung Kim, Ashraf Alattar, Jong Whan Jang, PaiChai University, Republic of Korea*

Concavities in the boundary of an object pose a challenge to active contour (snake) methods. In this paper, we present a snake-based scheme for efficiently detecting contours of objects with boundary concavities. The proposed method is composed of two steps. First, the object's boundary is detected using the proposed snake model. Second, snake points are optimized by inserting new points and deleting unnecessary points to better describe the object's boundary. The proposed algorithm can successfully extract objects with boundary concavities, and is insensitive to the number of initial snake points. Experimental results have shown that our algorithm produces more accurate segmentation results than the conventional algorithm.

11:30 - 11:50

**MA2-L1.3 ENTROPY AND MEMORY CONSTRAINED VECTOR QUANTIZATION WITH SEPARABILITY BASED FEATURE SELECTION**

*Sangho Yoon, Robert M. Gray, Stanford University, United States*

An iterative model selection algorithm is proposed. The algorithm seeks relevant features and an optimal number of codewords (or codebook size) as part of the optimization. We use a well-known separability measure to perform feature selection, and we use a Lagrangian with entropy and codebook size constraints to find the optimal number of codewords. We add two model selection steps to the quantization process: one for feature selection and the other for choosing the number of clusters. Once relevant and irrelevant features are identified, we also estimate the probability density function of irrelevant features in stead of discarding them. This can avoid the bias of problem of the separability measure favoring high dimensional spaces.

11:50 - 12:10

**MA2-L1.4 VIDEO OBJECT SEGMENTATION BASED ON OBJECT ENHANCEMENT AND REGION MERGING**

*Ken Ryan, Aishy Amer, Concordia University, Canada; Langis Gagnon, Computer Research Institute of Montreal (CRIM), Canada*

This paper proposes a number of improvements to existing work in off line video object segmentation. Object color and motion variance, and histogram-based merging are used to improve the initial segmentation. Segmentation quality measures taken from throughout the clip are used to enhance video objects. Cumulative histogram-based merging, occlusion handling, and island detection are used to help group regions into meaningful objects. Objective and subjective tests were performed on a set of standard video test sequences which demonstrate improved accuracy and greater success in identifying the real objects in a video clip compared to the reference method.



**MA2-L2 Processors & Multimedia II (Special Session)**

Time: Monday, July 10, 10:50 - 12:10

Place: Johnston

Co-Chairs: Yen-Kuang Chen, Intel Corporation and Eric Debes, Intel Corporation

10:50 - 11:10

**MA2-L2.1 SCCS: A SCALABLE CLUSTERED CAMERA SYSTEM FOR MULTIPLE OBJECT TRACKING COMMUNICATING VIA MESSAGE PASSING INTERFACE***Senem Velipasalar, Jason Schlessman, Cheng-Yao Chen, Wayne Wolf, Jaswinder Pal Singh, Princeton University, United States*

We introduce the Scalable Clustered Camera System, a peer-to-peer multi-camera system for multi-object tracking, where different CPUs are used to process inputs from distinct cameras. Instead of transferring control of tracking jobs from one camera to another, each camera in our system performs its own tracking and keeps its own tracks for each target object, thus providing fault tolerance. A fast and robust tracking method is proposed to perform tracking on each camera view, while maintaining consistent labeling. In addition, we introduce a new communication protocol, where the decisions about when and with whom to communicate are made such that frequency and size of transmitted messages are minimized. This protocol incorporates variable synchronization capabilities, so as to allow flexibility with accuracy tradeoffs. We discuss our implementation, consisting of a parallel computing cluster, with communication between the cameras performed by MPI. We present experimental results which demonstrate the success of the proposed peer-to-peer multi-camera tracking system, with accuracy of 95% for a high frequency of synchronization, as well as a worst-case of 15 frames of latency in recovering correct labels at low synchronization frequencies.

11:10 - 11:30

**MA2-L2.2 POWER-SCALABLE ALGORITHM AND RECONFIGURABLE MACRO-BLOCK PIPELINING ARCHITECTURE OF H.264 ENCODER FOR MOBILE APPLICATION***Yu-Han Chen, Tung-Chien Chen, Liang-Gee Chen, National Taiwan University, Taiwan*

In this paper, a power-scalable H.264 encoding system is provided with the efforts on both the algorithm and the architecture levels. For a start, a Motion Estimation (ME) pre-skip algorithm is adopted as a system-level power-scalable algorithm. In order to realize a dedicated hardware, a novel reconfigurable Macro-Block (MB) pipelining architecture is proposed. It can improve not only system flexibility but also hardware efficiency. Besides, it is also beneficial for power management with module-level gated clock insertion. According to simulation results, the proposed H.264 encoder can support power-scalable functionality in the range of about 20 to 90 mW with graceful quality degradation.

11:30 - 11:50

**MA2-L2.3 EMBEDDED CONVOLUTIONAL FACE FINDER***Sébastien Roux, Franck Mamalet, Christophe Garcia, France Telecom R&D division, France*

In this paper, we propose an optimization methodology and its application to implement the well-known Convolutional Face Finder (CFF) algorithm for real-time applications on cellular phone, such as advanced user interfaces, pictures indexing and security access control. The CFF is a feature extraction and classification technique which consists in a pipeline of convolutions and subsampling operations. Design of embedded systems must find a good trade off between performance and code size due to the limited amount of resource available. We propose a methodology to cope with the main drawbacks of the original implementation of CFF like floating-point computation and memory allocation, to allow parallelism exploitation and perform algorithm optimizations. Our embedded face detection system can accurately locate faces, and runs on a 275MHz Starcore DSP at 9 QCIF fr/s.

11:50 - 12:10

**MA2-L2.4 PAC DSP CORE AND APPLICATION PROCESSORS***David Chih-Wei Chang, I-Tao Liao, Industrial Technology Research Institute, Taiwan; Jenq-Kuen Lee, National Tsing Hua University, Taiwan; Wen-Feng Chen, Shau-Yin Tseng, Chien-Wei Jen, Industrial Technology Research Institute, Taiwan*

This paper provides an overview of the Parallel Architecture Core (PAC) project led by SoC Technology Center of Industrial Technology Research Institute (STC/ITRI) in Taiwan. The background of PAC project, a brief introduction to PAC core technologies, PAC SoC development suite, PAC benchmarks, and applications are presented. The main objective of the PAC development plan is to enhance Taiwan's competitiveness in the core technology related to key components, especially for portable multimedia applications.

**MA2-L3 Content Understanding II (Lecture)**

Time: Monday, July 10, 10:50 - 12:10

Place: Thomson

Chair: Jan Biemond, TU Delft

10:50 - 11:10

**MA2-L3.1 TOWARD INTELLIGENT USE OF SEMANTIC INFORMATION ON SUBSPACE DISCOVERY FOR IMAGE RETRIEVAL**
*Jie Yu, Qi Tian, University of Texas, San Antonio, United States*

Image retrieval has been widely used in many fields of science and engineering. The semantic concept of user interest is obtained by a learning process. Traditional techniques often assume the images are from certain distribution and all images from the same class are visually similar. Our study shows that those assumptions are inappropriate in many cases. To solve this problem we model the images as lying on non-linear subspaces embedded in the high dimensional space. We also find that a set of low-level feature subspaces may correspond to one high-level semantic concept. Unlike most unsupervised subspace learning techniques, we propose to intelligently use the semantic similarity and dissimilarity information provided by user in discovering the discriminant structure of image subspaces in respect to classification. Theoretical study shows that our methods converge to Linear Discriminant Analysis if certain criteria are met. Extensive experiments are designed to evaluate the performance of our method and compare it to other state-of-the-art techniques. The results show the superior performance of our proposed method.

11:10 - 11:30

**MA2-L3.2 PATTERN MINING IN VISUAL CONCEPT STREAMS**
*Lexing Xie, IBM Research, United States; Shih-Fu Chang, Columbia University, United States*

Pattern mining algorithms are often much easier applied than quantitatively assessed. In this paper we address the pattern evaluation problem by looking at both the capability of models and the difficulty of target concepts. We use four different data mining models: frequent itemset mining, k-means clustering, hidden Markov model, and hierarchical hidden Markov model to mine 39 concept streams from the a 137-video broadcast news collection from TRECVID-2005. We hypothesize that the discovered patterns can reveal semantics beyond the input space, and thus evaluate the patterns against a much larger concept space containing 192 concepts defined by LSCOM. Results show that HHMM has the best average prediction among all models, however different models seem to excel in different concepts depending on the concept prior and the ontological relationship. Results also show that the majority of the target concepts are better predicted with temporal or combination hypotheses, and there are novel concepts found that are not part of the original lexicon. This paper presents the first effort on temporal pattern mining in the large concept space. There are many promising directions to use concept mining to help construct better concept detectors or to guide the design of multimedia ontology.

11:30 - 11:50

**MA2-L3.3 MINING RELATIONSHIP BETWEEN VIDEO CONCEPTS USING PROBABILISTIC GRAPHICAL MODELS**
*Rong Yan, Ming-yu Chen, Alexander Hauptmann, Carnegie Mellon University, United States*

For large scale automatic semantic video characterization, it is necessary to learn and model a large number of semantic concepts. These semantic concepts do not exist in isolation to each other and exploiting this relationship between multiple video concepts could be a useful source to improve the concept detection accuracy. In this paper, we describe various multi-concept relational learning approaches via a unified probabilistic graphical model representation and propose using numerous graphical models to mine the relationship between video concepts that have not been applied before. Their performances in video semantic concept detection are evaluated and compared on two TRECVID'05 video collections.



11:50 - 12:10

**MA2-L3.4 AUTOMATIC SEMANTIC ANNOTATION OF IMAGES USING SPATIAL HIDDEN MARKOV MODEL**

*Feiyang Yu, Horace H. S. Ip, City University of Hong Kong, Hong Kong SAR of China*

This paper presents a new spatial-HMM(SHMM)for automatically classifying and annotating natural images. Our model is a 2D generalization of the traditional HMM in the sense that both vertical and horizontal transitions between hidden states are taken into consideration. The three basic problems with HMM-liked model are also solved in our model. Given a sequence of visual features, our model automatically derives annotations from keywords associated with the most appropriate concept class, and without pre-defining the keyword length threshold. Our experiments showed that our model outperformed the previous 2D MHMM in recognition accuracy and also achieved a high annotation accuracy.

**MA2-L4: Monday, July 10, 10:50 - 12:10**

**MA2-L4 Transcoding X to H.264 II (Special Session)**

Time: Monday, July 10, 10:50 - 12:10

Place: Toronto I

Chair: Hari Kalva, Florida Atlantic University

10:50 - 11:10

**MA2-L4.1 RD-OPTIMIZATION FOR MPEG-2 TO H.264 TRANSCODING**

*Gerardo Fernandez-Escribano, Universidad de Castilla-La Mancha, Spain; Hari Kalva, Florida Atlantic University, United States; Pedro Cuenca, Luis Orozco-Barbosa, Universidad de Castilla-La Mancha, Spain*

In this paper, we introduce and evaluate a low complexity macroblock partition mode decision algorithm for inter-frame prediction in MPEG-2 to H.264 transcoder. The proposed tools are used to compute an optimal MB coding mode decision with significantly reduced computational complexity. Specifically, we achieve the computational savings by using the following MB information coming from MPEG-2: the MB coding modes, the coded block pattern (CBPC) in MPEG-2, and the mean and variance of the 16 4x4 sub blocks of the MPEG-2 residual MBs. We use data mining algorithms to develop a decision tree for H.264 coding mode decisions. The decision trees are built using RD optimized mode decisions and result in highly efficient mode decisions. The proposed transcoder is 35% faster than the RD optimized H.264 reference transcoder without a significant PSNR degradation. The proposed transcoder performs over 3 dB better than the SAE cost based H.264 transcoding.

11:10 - 11:30

**MA2-L4.2 MOTION AND MODE MAPPING FOR MPEG-2 TO H.264/AVC TRANSCODING**

*Jun Xin, Anthony Vetro, Mitsubishi Electric Research Laboratories, United States; Shun-ichi Sekiguchi, Kazuo Sugimoto, Mitsubishi Electric Corporation, Japan*

This paper describes novel low-complexity transcoding techniques for an efficient MPEG-2 to H.264/AVC transcoder. We present a motion and mode mapping algorithm that directly maps incoming MPEG-2 motion information to H.264/AVC motion vectors as well as coding modes. Refinement strategies are also presented to improve performance. Experiment results demonstrate the effectiveness of the proposed low-complexity transcoder.

11:30 - 11:50

**MA2-L4.3 OPTIMAL REQUANTIZATION-BASED RATE ADAPTATION FOR H.264**

*Bo Shen, Hewlett-Packard Laboratories, United States*

Bit rate adaptation is one of the most important types of video transcoding. With H.264 becoming the predominant video codec of choice in video coding and streaming, prudent rate-adaptation techniques should be developed. In this paper, we investigate certain critical points in the spectrum of rate shaping requests. We show that the selection of quantization step sizes may not have monotonic effects on rate-distortion characteristics in the transcoding sense. In other words, distortion in the regular sense can be different from the distortion in the context of transcoding in which a requantization process is carried out. We show in a generic form that careful selections of the step size can lead to much improved performance. Experiments based on both simulation and real transcoding show the effectiveness of the proposed solution.

11:50 - 12:10

**MA2-L4.4 AN NEW COEFFICIENTS TRANSFORM MATRIX FOR THE TRANSFORM DOMAIN MPEG-2 TO H.264/AVC TRANSCODING**

*Gao Chen, Shouxun Lin, Yongdong Zhang, Gang Cao, Chinese Academy of Sciences, China*

In this paper, a fast transform method is proposed to convert MPEG-2 8-tap discrete cosine transform (DCT) coefficients to H.264/AVC 4-tap integer transform coefficients directly in the transform domain. The proposed transform method saves 16 operations for each 8x8 DCT block by utilizing a novel transform kernel matrix and a fast computing method for multiplication of this new matrix. The simulation results show that the proposed method causes only a very little quality degradation, which is completely negligible in practice with the maximum value lower than 8x10-3dB, as compared with Jun Xin's method. Hence, it can be efficiently used in the transform-domain MPEG-2 to H.264 transcoding.

**MA2-L5: Monday, July 10, 10:50 - 12:10**

**MA2-L5 Multimedia Interfaces II (Lecture)**

Time: Monday, July 10, 10:50 - 12:10

Place: Toronto II

Chair: Qiong Liu, FX Palo Alto Laboratory

10:50 - 11:10

**MA2-L5.1 BIOLOGICALLY VS. LOGIC INSPIRED ENCODING OF FACIAL ACTIONS AND EMOTIONS IN VIDEO**

*Michel Valstar, Maja Pantic, Imperial College London, United Kingdom*

Work on automatic facial expression analysis can roughly be divided into the recognition of complete facial expressions such as the six basic emotional states and the recognition of atomic facial muscle actions (Action Units, AUs). Classic psychological studies suggest that humans map AUs onto the basic emotion categories consciously using a finite number of rules while recent studies suggest that humans recognize emotions unconsciously with a process that is perhaps best modeled by artificial neural networks (ANNs). This paper investigates these issues. The results suggests that the two-step approach is possible with a small loss of accuracy and that biologically inspired classification techniques outperform those that approach the classification problem from a logical perspective, suggesting that the former are more suitable for computer-based analysis.

11:10 - 11:30

**MA2-L5.2 EVALUATION OF SELF-EDITING BASED ON BEHAVIORS-FOR-ATTENTION FOR DESKTOP MANIPULATION VIDEOS**

*Motoyuki Ozeki, Yuichi Nakamura, Kyoto University, Japan*

In this paper, we discuss a user interface issue with regard to automatic video editing based on the speaker's intentions. In our experiments, the subjects used previously developed video capturing system to employ 4 types of editing methods by making 3 types of presentations. Subjective evaluation revealed that the editing method that used behaviors-for-attention obtained a good score for a presentation in which the subject was provided with specific instructions regarding the tasks to be performed. In the case of a presentation without a scenario, an editing method using a footswitch and a posture obtained a higher score. It can be concluded that a combination of both behavior-based and footswitch-based editing would provide a good environment for content acquisition.

11:30 - 11:50

**MA2-L5.3 BRINGING THE WIKI COLLABORATION MODEL TO THE TABLETOP WORLD**

*Stefano Baraldi, Alberto Del Bimbo, Alessandro Valli, Università di Firenze, Italy*

We present an interactive workspace that integrates wiki collaboration in knowledge-building activities with face-to-face scenarios like brainstorming or problem solving sessions. A wiki serves as the repository for knowledge elements, which are presented to co-located users in form of a concept map projected on a table. A computer vision module tracks multiple hands and fingers on the table surface and map elements can be manipulated using a simple gesture language. Also smart devices, like PDAs or tablets, can be connected to the system and participate in the interaction with their local input. The concept map and the wiki are synchronized in real-time, providing notifications to both co-located and distributed users and allowing a community shared awareness that enhances and enriches the knowledge building experience.

11:50 - 12:10

**MA2-L5.4 DESIGNING INTERACTIONS IN EVENT-BASED UNIFIED MANAGEMENT OF PERSONAL MULTIMEDIA INFORMATION**

*Juan Pinzón, Rahul Singh, San Francisco State University, United States; Wolfgang Taube, Fachhochschule Furtwangen, Germany; Jensen Galan, San Francisco State University, United States*

Advancements in sensor technologies have made it easier and increasingly common to capture information using multiple media. This is especially true for personal multimedia information. Effective assimilation of such information requires recognizing the semantic correlations between media and the ability to model and interact with them in a unified manner. This paper presents our research in designing capabilities to support user-data interactions in context of the aforementioned issues. Central to our approach is characterization and modeling of media using the notion of an “event”. Building on this idea, we propose the design of operations as well as visualizations that not only allow event generation and manipulation, but also the ability to interact with various semantically important characteristics of the underlying information. Experimental and comparative evaluations demonstrate the efficacy and promise of the approach.

**MA2-L6: Monday, July 10, 10:50 - 12:10**

**MA2-L6 Multimedia Database Applications II (Lecture)**

Time: Monday, July 10, 10:50 - 12:10

Place: Varley

Chair: Dan Schonfeld, University of Illinois, Chicago

10:50 - 11:10

**MA2-L6.1 TENSOR-BASED MULTIPLE OBJECT TRAJECTORY INDEXING AND RETRIEVAL**

*Xiang Ma, Faisal Bashir, Ashfaq Khokhar, Dan Schonfeld, University of Illinois at Chicago, United States*

This paper presents novel tensor-based object trajectory modelling techniques for content-based indexing and retrieval of multiple objects in video data. Three different tensor decomposition techniques-PARAFAC, HOSVD and Multiple-SVD are explored to achieve the goal of representing multiple trajectory data using a minimum set of coefficients and data-dependant bases. Both full-trajectory and segmented-trajectory tensors are tested. Our simulation results show that the PARAFAC-based approach provides superior precision-recall metrics, smaller query processing time and higher compression ratio compared to other tensor-based approaches.

11:10 - 11:30

**MA2-L6.2 POSTER IMAGE MATCHING BY COLOR SCHEME AND LAYOUT INFORMATION**

*Cheng-Yao Chen, Takayuki Kurozumi, Junji Yamato, NTT Communication Science Laboratories, Japan*

In this paper, we demonstrate a novel poster image matching system for wireless multimedia applications. We propose a method that incorporates both color and layout information of the poster image to achieve a robust performance in poster image matching. We apply both color compensation and background separation to extract a poster from an image effectively. Based on our experiment, we show that even under the effects of lighting, image rotation, scaling, and occlusion, our system can still maintain high recall and precision. We also show that our system can recognize the correct image from a database which contains several poster images with similar features. Finally, the promising performance of the poster image matching encourages us to further enrich the information retrieval for wireless environment.

11:30 - 11:50

**MA2-L6.3 IMPROVED SIMILARITY-BASED ONLINE FEATURE SELECTION IN REGION-BASED IMAGE RETRIEVAL**

*Fei Li, Qionghai Dai, Wenli Xu, Tsinghua University, China*

To bridge the gap between high level semantic concepts and low level visual features in content-based image retrieval (CBIR), online feature selection is really required. An effective similarity-based online feature selection algorithm in region-based image retrieval (RBIR) systems was proposed by W. Jiang etc., but some parts of the algorithm need to be improved. In this paper, the above algorithm is modified in two aspects: (1) Adaptive mixture models based on mutual information theory are adopted to determine the codebook size. (2) A new method is proposed, which can select not only feature axes parallel to the original ones, but also combined feature axes. Experimental results on 10000 images show that the proposed method can improve the retrieval performance, and save the computational time.

**MA2-L6.4 LARGE-SCALE DUPLICATE DETECTION FOR WEB IMAGE SEARCH**

*Bin Wang, University of Science and Technology of China, China; Zhiwei Li, Mingjing Li, Wei-Ying Ma, Microsoft Research Asia, China*

Finding visually identical images in large image collections is important for many applications such as intelligence propriety protection and search result presentation. Several algorithms have been reported in the literature, but they are not suitable for large image collections. In this paper, a novel algorithm is proposed to handle the situation, in which each image is compactly represented by a hash code. To detect duplicate images, only the hash codes are required. In addition, a very efficient search method is implemented to quickly group images with similar hash codes for fast detection. The experiments show that our algorithm can be both efficient and effective for duplicate detection in web image search.

### MA2-P1: Monday, July 10, 10:50 - 11:50

**MA2-P1 VLSI System-on-Chip Integrated Multimedia Systems (Poster)**

Time: Monday, July 10, 10:50 - 11:50

Place: Toronto III

Chair: Liang-Gee Chen, National Taiwan University

**MA2-P1.1 A HIGH THROUGHPUT VLSI ARCHITECTURE DESIGN FOR H.264 CONTEXT-BASED ADAPTIVE BINARY ARITHMETIC DECODING WITH LOOK AHEAD PARSING**

*Yao-Chang Yang, Chien-Chang Lin, Hsui-Cheng Chang, National Chung Cheng University, Taiwan; Ching-Lung Su, National Yunlin University of Science and Technology, Taiwan; Jiun-In Guo, National Chung Cheng University, Taiwan*

In this paper we present a high throughput VLSI architecture design for Context-based Adaptive Binary Arithmetic Decoding (CABAD) in MPEG-4 AVC/H.264. To speed-up the inherent sequential operations in CABAD, we break down the processing bottleneck by proposing a look-ahead codeword parsing technique on the segmenting context tables with cache registers, which averagely reduces up to 53% of cycle count. Based on a 0.18 micrometer CMOS technology, the proposed design outperforms the existing design by both reducing 40% of hardware cost and achieving about 1.6 times data throughput at the same time.

**MA2-P1.2 AN EFFICIENT REFERENCE FRAME STORAGE SCHEME FOR H.264 HDTV DECODER**

*Peng Zhang, Wen Gao, Di Wu, Chinese Academy of Sciences, China; Don Xie, Grandview Semiconductor, China*

This paper proposes an efficient reference frame storage scheme for HDTV VLSI decoder to reduce external memory bandwidth requirement. The proposed scheme consists of the pixel duplication mechanism and the L-C (luma-chroma) correlated mapping method. Pixel duplication completely eliminates the possibility of an access crossing word boundary and therefore substantially increases the memory bandwidth efficiency. L-C correlated mapping exploits address relationships between the luma and chroma reference pixels and largely reduces bank conflict overhead of memory accesses. The two mechanisms combined together efficiently improve the bandwidth usage: up to 47% bandwidth in worst case is saved compared with the previous schemes, and 25% in average case.

**MA2-P1.3 SCALABLE RATE-DISTORTION-COMPUTATION HARDWARE ACCELERATOR FOR MCTF AND ME**

*Yi-Hau Chen, Ching-Yeh Chen, Chih-Chi Cheng, Liang-Gee Chen, National Taiwan University, Taiwan*

Motion-Compensated Temporal Filtering (MCTF) is an innovative prediction scheme for video coding and the core technology of scalable extension of H.264/AVC. The first MCTF and ME hardware work is in this paper. The proposed hardware not only can support the various coding schemes in JSVM and H.264 but also can adapt itself to provide rate-distortion-computation scalability. With the frame-level searching range data reuse and frame-interleaved MB pipelining scheme, external memory bandwidth are reduced 33%, and 10 Kbits buffer are saved. The proposed MCTF/ME hardware is of 350K gate counts and 30KB internal buffer, which can perform four-level MCTF or H.264 P-/B-frames at CIF Format.

#### **MA2-P1.4 A NOVEL DATA-PARALLEL COPROCESSOR FOR MULTIMEDIA SIGNAL PROCESSING**

*Lai Mingche, Dai Kui, Lu Hongyi, Wang Zhiying, National University of Defense Technology, China*

Realizations of demanding embedded applications particularly in the field of signal processing often require high processing performance and low energy consumption which are far beyond what can be delivered by DSPs nowadays. A novel data-parallel coprocessor basing on the Transport Triggered Architecture (TTA) is presented in this paper. The coprocessor consists of two powerful arithmetic clusters, the stream memory as well as the optimized data transport network, and is good at exploiting the data parallelism in the computation intensive multimedia applications. Also, the solution of the clock-gating in a gradual way is advanced for the low power dissipation. Then, a SoC chip involving the coprocessor is implemented using 0.18 $\mu$ m CMOS process. Experimental results show that this coprocessor has good performance improvement for multimedia applications.

### **MA2-P2: Monday, July 10, 10:50 - 11:50**

#### **MA2-P2 Multimedia Interfaces: Searching and Browsing (Poster)**

Time: Monday, July 10, 10:50 - 11:50

Place: Toronto III

Chair: Hidenobu Nagata, NTT Corporation

#### **MA2-P2.5 VIDEO POT: INDEXING-BASED DESKTOP VIDEO SEARCH SYSTEM**

*Hidenobu Nagata, Dan Mikami, Shozo Azuma, Masashi Morimoto, NTT Corporation, Japan*

We present VideoPot, a desktop video search system designed for digital video files stored on personal computers. The core metric of this system is based on video-indexing technology and an automatic related-term collection framework. Digital files on local disks are crawled, and images of representative scenes are extracted from video files using video-indexing technology. Related terms, used as target information in the video search, are collected from both local files and Web pages, using an automatic related-term collection algorithm. These images and terms are then stored in the database and associated with the original videos. We describe a unique search result interface that helps users distinguish the target video. In the search result interface, the results of the video search are listed at two levels to search videos faster. The effectiveness of images on our interface derived from video-indexing is shown by a comparison with a conventional file search application.

#### **MA2-P2.6 ADVANCING CONTENT-BASED RETRIEVAL EFFECTIVENESS WITH CLUSTER-TEMPORAL BROWSING IN MULTILINGUAL VIDEO DATABASES**

*Mika Rautiainen, Tapio Seppänen, Timo Ojala, University of Oulu, Finland*

Interactive experiments on video retrieval systems need to address the problem of internal validity, i.e. how much the test users' experience affects the retrieval effectiveness. This paper compares the semantic retrieval performance of novice users and expert system developers. The test system utilizes cluster-temporal browsing, which combines chronological video structure and computation of similarities into single interface. Interactive experiments with eight test users were carried out in a database of ~80 hours of multilingual news video from TRECVID 2005 benchmark. A cluster-temporal browser was found to improve the retrieval effectiveness by 12% with novice system users. Expert users were able to achieve 18% better performance than the novice users. Additionally, manual search experiments demonstrated that search performance can be improved by 19-25% when a plain text search is supplemented with content-based features.

#### **MA2-P2.7 PREDICTIVE DYNAMIC USER INTERFACES FOR INTERACTIVE VISUAL SEARCH**

*Sam Mavandadi, Parham Aarabi, Azadeh Khaleghi, Ron Appel, University of Toronto, Canada*

This paper proposes a method for designing user interfaces based on ideas rooted in data communication theory. It suggests that a visual user interface should be treated as a multi-transmitter, single-receiver communication system, where the total available bandwidth for transmission is limited. The proposed design entails the scaling of visual components that are displayed according to their degree of relevance to the user, or in other words, their probability of selection by the user.

## **MA2-P2.8 USABILITY EVALUATION FOR IMAGE RETRIEVAL BEYOND DESKTOP APPLICATIONS**

*Thomas Käster, Vis-à-pix, Germany; Michael Pfeiffer, Bielefeld University, Germany; Christian Bauckhage, Telekom AG Laboratories, Germany*

Interactivity is a key concept in modern content-based retrieval. Therefore, in addition to the ability to learn from user generated data, easy and intuitive to use interfaces are an important area of research in (multi)media retrieval. In this contribution, we focus on the latter aspect and present how different modalities like speech and gestures on super sized touch screen facilities may be integrated to accomplish the goal of intuitive interaction. In order to evaluate our approach, we conducted a series of usability experiments. Their results demonstrate that our multimodal user interface allows for both, comfortable and successful interactive image retrieval.

## **MA2-P2.9 SHAPE-BASED VISUAL QUERY REWRITING**

*Georges Chalhoub, Richard Chbeir, Kokou Yetongnon, Université de Bourgogne - Le2i, France*

A visual query is based on pictorial representation of conceptual entities and operations. One of the most important features used in visual queries is the shape. Despite its intuitive writing, a shape-based visual query usually suffers of a complexity processing related to two major parameters: 1-the imprecise user request, 2-shapes may undergo several types of transformation. Several methods are provided in the literature to assist the user during query writing. In this paper, we present a new cooperative approach based on the shape neighborhood concept allowing the user to rewrite a shape-based visual query according to his preferences with high flexibility in terms of including (or excluding) only some shape transformations and of result sorting

## **MA2-P2.10 MINING TEXT AND VISUAL LINKS TO BROWSE TV PROGRAMS IN A WEB-LIKE WAY**

*Xin Fan, University of Science and Technology of China, China; Hisashi Miyamori, National Institute of Information and Communications Technology, Japan; Katsumi Tanaka, National Institute of Information and Communications Technology / Kyoto University, Japan; Mingjing Li, Microsoft Research Asia, China*

As the amount of recoded TV content is increasing rapidly, people need active and interactive browsing methods. In this paper, we use both text information from closed captions and visual information from video frames to generate links to enable one to explore not only the original video content but also augmented information from the Web. This solution especially shows its superiority when the video content cannot be well represented only by closed captions. A prototype system was implemented and some experiments were carried out to prove the effectiveness and efficiency.

## **MA2-P2.11 TOWARDS AN EFFICIENT INTEGRATION, STRUCTURE AND EXPLORATION OF LANDSCAPE ARCHITECTURE PROJECT INFORMATION**

*Franck Favetta, Ecole Nationale Supérieure de la Nature et du Paysage, France; Robert Laurini, Laboratoire d'InfoRmatique en Images et Systemes d'information, France*

Landscape Architecture projects have many specific requirements such as particular multimedia and geographic data integration and structure, information preview, user-friendly interface, and means of multi-actor participation. This article presents a solution for an efficient, quick and user-friendly integration, structure, exploration and management of landscape information. Our proposal extends different existing solutions and introduces useful preview abilities. A recently developed prototype implements the solution.

**MA2-P3 Multimedia Interfaces III (Poster)**

Time: Monday, July 10, 10:50 - 11:50

Place: Toronto III

Chair: Jar-Ferr Yang, National Cheng Kung University

**MA2-P3.12 AUTOMATIC GEOMETRIC AND PHOTOMETRIC CALIBRATION FOR TILING MULTIPLE PROJECTORS WITH A PAN-TILT-ZOOM CAMERA**

*Yu-Pao Tsai, National Chiao Tung University, Taiwan; Yen-Nien Wu, Shou-Chun Liao, National Taiwan University, Taiwan; Zen-Chung Shih, National Chiao Tung University, Taiwan; Yi-Ping Hung, National Taiwan University, Taiwan*

Research on creating a large, high-resolution, low-cost display system has become increasingly important due to the growing desire in many fields for bigger and better displays. The goal of this research is to build a seamless large-scale display system by tiling multiple projectors with the help of a pan-tilt-zoom camera. In order to achieve this goal, we took two tasks into consideration: geometric calibration and photometric calibration. Compared to the previous work, our method for geometric calibration is more accurate, thanks to the much higher resolution images acquired by combining several zoom-in images and doing lens correction for the camera at first. Expensive equipments such as spectroradiometer and colorimeter are commonly used for photometric calibration. With our method, a comparatively low-cost camera can be used instead, because we adopt the technique of producing a high dynamic range image from several images of different exposures. In our experiments, the average error of photometric measurement achieved by our method is less than the mean perceptibility tolerance. Our method has great potential for many applications that require large and high-resolution displays.

**MA2-P3.13 VIDEO STABILIZATION PERFORMANCE ASSESSMENT**

*Matti Niskanen, Olli Silvén, University of Oulu, Finland; Marius Tico, Nokia Research Center, Finland*

Shooting videos with a hand-held camera introduces shaking, which incontrovertibly reduces video quality. Digital video stabilization is a process to compensate for camera motion by means of image processing. In the best case, it not only removes the image motion, but also reduces image distortion caused by unintentional camera motion. In practice, removing solely unwanted jitter cannot be achieved precisely. Furthermore, the stabilization process itself often introduces some additional distortion in images instead of removing it. In this paper, various means to automatically evaluate the performance of the video stabilization process are proposed, based on measuring the divergence and jitter of the remaining unintentional motion and blurring using point spread function (PSF). This helps, for example, in tuning the system parameters for better quality.

**MA2-P3.14 THE HYDRAULOPHONE: INSTRUMENTATION FOR TACTILE FEEDBACK FROM FLUID STREAMS AS A NEW MULTIMEDIA INTERFACE (WATER FOUNTAIN AS KEYBOARD)**

*Steve Mann, Ahmedullah Sharifi, Mike Hung, Russell Verbeeten, University of Toronto, Canada*

Water fountains are proposed as haptic/tactile surfaces that respond when people touch, block, or restrict flow of water jets to trigger multimedia events. This "fluid-user-interface" metaphor is based on the "hydraulophone" (alternative spelling: "hydrauliphone"), a musical instrument played by direct interaction with a pressurized hydraulic fluid emerging from an array of finger holes. The hydraulophone is like a keyboard, made from a water fountain, but each jet is a soft key that can be pressed in infinitely many different ways to obtain intricate and independent control of the volume, pitch, and timbre of each note, when sounding multiple notes simultaneously (i.e. playing harmelody). Some "pneumatophones" (alternative spelling: "pneumatiphone") have been designed to work on air. In their purely acoustic form, pneumatophones are woodwind instruments similar to flutes except that they can play chords, and there is a rich space of variations in tone, volume, timbre, etc., that can be independently applied to each note when more than one note is sounded together. This paper describes the construction of a MIDI interface that can be attached to an acoustic hydraulophone, or pneumatophone, or to an ordinary water fountain, or air fountain. This allows the fountain to be used as a haptic/tactile control surface or highly expressive fluid keyboard that provides gentle, soothing (soft) tactile feedback quite different from either the hard (solid) feedback of plastic keys, or the total lack of tactile feedback inherent in camera-based, vision-based, or proximity-sensing user-interfaces.



## **MA2-P3.15 SUBJECTIVE EVALUATIONS OF AN EXPERIMENTAL GESTUREPHONE**

*Mohd Nazri Ramliy, Nor Azhar Mohd Arif, Ryoichi Komiya, Multimedia University, Malaysia*

This paper presents the findings of our subjective evaluations on the integration of gestures in telecommunication. The experimental setup for tracking and imitating the human arm gesture are described. Our research investigates the possibility of transferring this often overlooked communication medium in our daily communication, for its application in telecommunication using robotics. Based on the subjective evaluation, a maximum allowable delay for an imperceptible gesture reconstruction in the lateral setup is suggested.

## **MA2-P3.16 AUTOFRAMING: A RECOMMENDATION SYSTEM FOR DETECTING UNDESIRABLE ELEMENTS AND CROPPING AUTOMATICALLY PHOTOS**

*Frank Nielsen, Shigeru Owada, Sony Computer Science Laboratories, Inc., Japan; Yuichi Hasegawa, Sony Corporation, Japan*

In this paper, we present a recommendation system for automatically recentring and cropping digital still pictures that exhibit capturing artefacts. Autoframing images not only yields better visual pictures but more importantly allows us to remove undesirable artefacts such as lens obstructions by fingers, cellphone straps, or back heads. We report on our real-time prototype system that is targeted to consumer digital still cameras.

## **MA2-P3.17 AN INTELLIGENT GUIDING BULLETIN BOARD SYSTEM WITH REAL-TIME VISION AND MULTI-KEYWORD SPOTTING MULTIMEDIA HUMAN-COMPUTER INTERACTION**

*Cheng-Yu Chang, Chung-Hsien Yang, You-Sheng Yeh, Pau-Choo Chung, Jhing-Fa Wang, Jar-Ferr Yang, National Cheng Kung University, Taiwan*

This paper presents an intelligent guiding bulletin board system (iGBBS), which is based on vision-interactive and multiple keyword-spotting technology. The system is aimed to provide different kinds of multimedia human-computer interaction (MMHCI) for users under different requirements. At first, a real-time front-view face detection using Harr-like features is used to decide when iGBBS should wake up and become interactive with the user. After system initialization, some feature points within the detected face area are going to be found. Then the orientation of user's head will be estimated via pyramidal Lucas-Kanade optical flow tracking. In addition, spotting the keyword from user's utterance with some related augmented reality responses would be provided as well. The performance of vision-interaction in iGBBS could be reached to 20 fps under Pentium IV 1G Hz PC. The error rate of multiple keyword-spotting interaction in iGBBS is about 36.2% and people can get the right response in 2.76 times search averagely. With the comparison to the traditional guiding system, bulletin board, or other non-vision-based input devices system, such like gloves or markers, our system offers a simple, useful and economical solution for the real-time interaction between the user and computer.

## **MA2-P3.18 CONSISTENT GOAL-DIRECTED USER MODEL FOR REALISTIC MAN-MACHINE TASK-ORIENTED SPOKEN DIALOGUE SIMULATION**

*Olivier Pietquin, SUPELEC, France*

Because of the great variability of factors to take into account, designing a spoken dialogue system is still a tailoring task. Rapid design and reusability of previous work is made very difficult. For these reasons, the application of machine learning methods to dialogue strategy optimization has become a leading subject of researches this last decade. Yet, techniques such as reinforcement learning are very demanding in training data while obtaining a substantial amount of data in the particular case of spoken dialogues is time-consuming and therefore expansive. In order to expand existing data sets, dialogue simulation techniques are becoming a standard solution. In this paper we describe a user modeling technique for realistic simulation of man-machine goal-directed spoken dialogues. This model, based on a stochastic description of man-machine communication, unlike previously proposed models, is consistent along the interaction according to its history and a predefined user goal.

## **MA2-P3.19 SIGN LANGUAGE RECOGNITION FROM HOMOGRAPHY**

*Qi Wang, Harbin Institute of Technology, China; Xilin Chen, Chunli Wang, Wen Gao, Chinese Academy of Sciences, China*

It is difficult to recognize sign language in different viewpoint. The HMM method is hindered by the difficulty of extracting view invariant features. The general template matching methods have a strong constraint such as accurate alignment between the template sign and the test sign. In the paper, we introduce a novel approach for viewpoint invariant sign language recognition. The proposed approach requires no view invariant features, low training and no alignment. Its basic idea is to consider a sign as a series of tiny hand motions and utilize the HOMOGRAPHY of tiny hand motions. Using the word of "homography", we mean that there are the same tiny hand motions as well as their appearance order in different performances of the same sign. The experimental results demonstrate the efficiency of the proposed method.



**MA2-P4 Emerging Technologies for Next-Generation Multimedia Systems: Challenges and Opportunities**  
(Poster)

Time: Monday, July 10, 10:50 - 11:50

Place: Toronto III

Chair: Hiroki Nakano, IBM Japan Ltd.

**MA2-P4.20 EXPERIENCES WITH HIGH DEFINITION INTERACTIVE VIDEO CONFERENCING**

*Ladan Gharai, Tom Lehman, University of Southern California, United States; Alvaro Saurin, Colin Perkins, University of Glasgow, United Kingdom*

We review the design and implementation of UltraGrid, a new high definition video conferencing system, and present some experimental results. UltraGrid was the first system to support gigabit rate high definition interactive video conferencing on commodity systems and networks, and we present measurements to illustrate behavior of production networks subject to such real time traffic. We illustrate the benefits of hybrid IP/provisioned optical networks over best effort IP networks for this class of traffic, and motivate the development of congestion control algorithms for interactive conferencing on best effort IP networks.

**MA2-P4.21 MULTIPOINT MEASURING SYSTEM FOR VIDEO AND SOUND - 100-CAMERA AND MICROPHONE SYSTEM -**

*Toshiaki Fujii, Kensaku Mori, Kazuya Takeda, Kenji Mase, Masayuki Tanimoto, Yasuhito Suenaga, Nagoya University, Japan*

We developed a novel multipoint measurement system capable of acquiring video and sound at more than 100 points in a “synchronized” manner. In this paper, we first describe the specification of the system and how the system works in detail. Then we report some experimental results that confirm the performance of the system. We also describe test data set we provided for MPEG(Moving Picture Experts Group) Multi-viewpoint Video Coding activities. Using this system, we are planning to conduct projects to measure humans and their activities, collect a large volume of real-world data of video and sound, and release them to the public.

**MA2-P4.22 A LANGUAGE AND ARCHITECTURE FOR AUTOMATING MULTIMEDIA CONTENT PRODUCTION ON GRID**

*Pierfrancesco Bellini, Ivan Bruno, Paolo Nesi, University of Florence, Italy*

Possible solutions to the management of multichannel delivering, production on demand, and containment of sale prices in the digital multimedia content production could be the automating, accelerating and restructuring the production process. The proposed solution provides innovative methods and tools to speed up and optimize content production and distribution based on the GRID technology supported by a specific programming language that allow defining the automatic procedure for content processing, production, adaptation, protection, DRM managing, distribution, etc. This paper describes the GRID architecture of the AXMEDIS Content Processing Area and the language adopted to define algorithms executed into the GRID environment.

**MA2-P4.23 CELL BROADBAND ENGINE BASED REALTIME WAVELET DECOMPOSITION FOR HDTV VIDEO IMAGE AND BEYOND**

*Akihiro Asahara, Munehiro Doi, Yumi Mori, Hiroki Nishiyama, Hiroki Nakano, IBM Japan Ltd., Japan*

The Cell broadband engine (CBE) is a novel multi-core microprocessor designed to provide compact and high-performance processing capabilities for a wide range of applications. Real-time image processing applications having parallelism with large amounts of data are good examples to demonstrate the unique capabilities of the CBE. In this paper, we describe the evaluation of the performance for image processing using wavelet transforms on CBE. Our results show that the CBE is extremely efficient in the processing performance compared with commercially available processors, and thus, we conclude that the CBE is quite suitable for next generation large pixel formats, such as 4K/2K-Digital Cinema.

**MA2-P5 Media Signal Processing I (Poster)**

Time: Monday, July 10, 10:50 - 11:50

Place: Toronto III

Chair: Dan Schonfeld, University of Illinois, Chicago

**MA2-P5.24 A NEW DEBLOCKING ALGORITHM BASED ON ADJUSTED CONTOURLET TRANSFORM**

*Haohao Song, Songyu Yu, Chen Wang, Li Song, Hongkai Xiong, Shanghai Jiao Tong University, China*

A new postprocessing method based on adjusted contourlet transform is introduced in this paper for suppressing blocking artifacts (BA) in block-based discrete cosine transform (BDCT) compressed images. To our best knowledge, this is the first time contourlet is applied to this field. By exploiting scale space edge detector (ss-edge detector), our algorithm can extract and protect blocking map (BM) and edge map (EM) in the compressed image respectively in the same time. By transforming the compressed image into adjusted contourlet domain, the adaptive thresholds are obtained according to BM. According to the adaptive thresholds, the contourlet coefficients in different subbands are filtered. Experimental results show that our deblocking algorithm achieves better performance than the other iterative and noniterative methods reported in the literature.

**MA2-P5.25 RESTORATION OF HALFTONED COLOR-QUANTIZED IMAGES USING LINEAR ESTIMATOR**

*Yik-Hing Fung, Yuk-Hee Chan, Hong Kong Polytechnic University, Hong Kong SAR of China*

Restoration of color-quantized images in which error diffusion is involved is rarely addressed in the literature. Conventional restoration algorithms were generally proposed for handling other noise models and hence they are not effective in restoring color-quantized images. In this paper, we tackle this problem by making use of the best linear estimation. It is shown by the simulation results that the algorithm can improve the quality of color-quantized images to a certain extent.

**MA2-P5.26 MODIFIED LAPLACIAN FILTER WITH INTENSITY CORRECTION TECHNIQUE FOR IMAGE RESOLUTION ENHANCEMENT**

*Day-Fann Shen, Chui-Wen Chiu, Pon-Jay Huang, National Yunlin University of Science and Technology, Taiwan*

By analyzing the relationship between a high-resolution image sensor array and the corresponding degraded low-resolution image, a modified 3x3 Laplacian filter (MLF) and the intensity correction technique (IC) are proposed for properly restoring the frequency components attenuated in the degradation process, thus reveal more detailed contents. Extensive experiments show that the proposed MLF outperforms several recently developed algorithms both in perceptual quality (especially in the texture areas) and in objective quality in terms of PSNR. Theoretically, the proposed SR algorithm can be applied to any digital image for increased resolutions; in addition, the simplicity of proposed algorithm is quite desirable for many real-world applications.

**MA2-P5.27 RAIN REMOVAL IN VIDEO BY COMBINING TEMPORAL AND CHROMATIC PROPERTIES**

*Xiaopeng Zhang, Hao Li, Yingyi Qi, Wee Kheng Leow, Teck Khim Ng, National University of Singapore, Singapore*

Removal of rain streaks in video is a challenging problem due to the random spatial distribution and fast motion of rain. This paper presents a new rain removal algorithm that incorporates both temporal and chromatic properties of rain in video. The temporal property states that an image pixel is never always covered by rain throughout the entire video. The chromatic property states that the changes of R, G, and B values of rainaffected pixels are approximately the same. By using both properties, the algorithm can detect and remove rain streaks in both stationary and dynamic scenes taken by stationary cameras. To handle videos taken by moving cameras, the video can be stabilized for rain removal, and destablized to restore camera motion after rain removal. It can handle both light rain and heavy rain conditions. Experimental results show that the algorithm performs better than existing algorithms.

## **MA2-P5.28 OPTIMAL LINEAR COMBINATION OF DENOISING SCHEMES FOR EFFICIENT REMOVAL OF IMAGE ARTIFACTS**

*Ramin Eslami, Hayder Radha, Michigan State University, United States*

Different denoising schemes show dissimilar types of artifacts. For example, certain transform-based denoising schemes could introduce artifacts in smooth regions while others eliminate texture regions. Using different schemes for denoising a noisy image, we can consider the denoised results as different estimates of the image. Through linear combination of the results, we minimize the  $l_2$  norm of the error to find the optimum coefficients in a least-square-error sense. We employ the wavelet transform, contourlet transform, and adaptive 2-D Wiener filtering as our denoising schemes. Then we apply the proposed method to the denoised results of the individual schemes. This approach eliminates most of the artifacts and achieves significant improvement in the PSNR values. We also propose averaging of the denoised results as a special case of linear combination and show that it yields near-optimal performance.

## **MA2-P5.29 CODING ARTIFACT REDUCTION USING NON-REFERENCE BLOCK GRID VISIBILITY MEASURE**

*Ihor Kirenko, Remco Muijs, Ling Shao, Philips Research Laboratories, Netherlands*

In this work a new method is proposed for coding artifact reduction of MPEG compressed video sequences. The method makes use of a simple cost-effective technique that allows the block grid position and its visibility to be determined without the need for access to the coding parameters. This information, combined with the results of local spatial analysis of luminance and chrominance components of a decoded image, is used to effectively suppress coding artifacts while preserving the sharpness of object edges. The results of our experiments show the high efficiency of the proposed approach.

## **MA2-P5.30 AN ADAPTIVE DE-INTERLACING ALGORITHM BASED ON TEXTURE AND MOTION VECTOR ANALYSIS**

*Jianguo Du, Songnan Li, Debin Zhao, Harbin Institute of Technology, China; Qian Huang, Wen Gao, Chinese Academy of Sciences, China*

In this paper, we propose a novel hybrid de-interlacing algorithm, which effectively combines two motion-compensated (MC) de-interlacing techniques: MC Median Filtering (MCMF) and Adaptive Recursive (AR) with one spatial approach: Line Averaging (LA). Despite its drawbacks, AR is one of the best methods nowadays. MCMF helps reduce flickers and LA is very robust to erroneous motion vectors. The interpolation switches among these methods based on the proposed measurement of texture smoothness and motion vector (MV) reliability. MCMF is adopted when MV is reliable and texture is rich. LA is used when MV is unreliable and texture is smooth. AR is applied to the remaining regions. Experimental results show that the proposed algorithm is superior to the compared algorithms in terms of peak signal-to-noise ratio (PSNR) and that the de-interlaced videos have very high subjective quality.

## **MA2-P5.31 DETECTION AND REMOVAL OF LONG SCRATCH LINES IN AGED FILMS**

*Timothy K. Shih, Louis H. Lin, Tamkang University, Taiwan; Wonjun Lee, Korea University, Republic of Korea*

Historical films usually have defects. We study the type of defects, and propose a series of solutions to detect defects before they are repaired by our inpainting algorithms. This paper focuses on a difficult issue to locate long vertical line defects in aged films. A progressive detection algorithm is proposed. We are able to detect more than 86% (recall rate) of effective line defects. These line defects are then removed step by step. The experiments use real historical video collected from national museum and public channel, instead of using computer generated noise. The results are visually pleasant based on our subjective evaluation by volunteers.

## **MA2-P5.32 POWER-AWARE PARTICLE FILTERING FOR VIDEO TRACKING**

*Pan Pan, Dan Schonfeld, University of Illinois at Chicago, United States*

This paper presents a novel approach to particle filtering which minimizes the total tracking distortion by considering dynamic variance of proposal density and adaptive number of particles for each frame. Traditionally, particle filters use fixed variance of proposal density and fixed number of particles per frame. We first propose the tracking distortion measurement and then obtain the optimal particle number and memory size allocation equations under two different constraints. After that, the optimal particle number allocation equation is demonstrated in one-dimensional and two-dimensional object tracking. Experimental results show the improved performance of our power-aware particle filters in comparison to traditional particle filters. At last, we give the complete algorithm for real application and show the better performance. To the best of our knowledge, this paper is the first to consider the variant numbers of particles for each frame.

### **MA2-P5.33 MODELING BLOCKING VISUAL SENSITIVITY PROFILE**

*Guangtao Zhai, Wenjun Zhang, Xiaokang Yang, Yi Xu, Shanghai Jiao Tong University, China*

Blocking artifact is the most prevailing degradation caused by block-based DCT coding techniques under low bit-rate conditions. To alleviate blockings perceptually, it is desirable to measure the visibility of blocking artifacts. In this paper, we propose an efficient method of estimating the visual sensitivity of blocking artifacts in block-based DCT coding. The differences on block boundaries are measured and transformed into block discontinuity map. We consider the effects of luminance adaptation and texture masking on the blockings and integrate them using nonlinear operator to form an overall masking map. This masking map is then incorporated with the discontinuity map to generate the Blocking Visual Sensitivity Map (BVSM). This map can be used to guide perceptual quality assessment, codec parameter optimization, post-processing, etc. We demonstrate the validity of the BVSM through its application in image quality assessment.

### **MA2-P5.34 DIRECTIONAL WEIGHTING-BASED DEMOSAICKING ALGORITHM FOR NOISY CFA ENVIRONMENTS**

*Hung-Yi Lo, Tsung-Nan Lin, Chih-Lung Hsu, Cheng-Hsien Lee, National Taiwan University, Taiwan*

Captured CFA data by image sensors like CCD/or CMOS are often corrupted by noises. To produce high quality images acquired by CCD/CMOS digital cameras, the problem of noise needs addressing. In this paper, we propose a novel demosaicking algorithm with the ability to handle noisy CFA data directly. By utilizing the proposed spatial filter which can characterize the similarity likelihood in local structure accurately, the noisy pixel is then filtered depending on the degree of similarity between the current pixel and a weighted average of its neighboring pixels. Therefore the edge information can be preserved without the blurring artifacts while the capacity of noise reduction can be adjusted to the maximum degree in the smooth region. Unlike traditional approaches which apply the noise reduction on the 24-bit color image, our algorithm is the first one that can accomplish the demosaicking processing and noise removal simultaneously, which contributes to the reduction of hardware cost since one module can achieve two functions efficiently at the same time.

## **MP1-L1: Monday, July 10, 13:10 - 14:30**

### **MP1-L1 Feature Extraction and Classification I (Lecture)**

Time: Monday, July 10, 13:10 - 14:30

Place: Carmichael

Chair: Irene Gu, Chalmers University of Technology

13:10 - 13:30

### **MP1-L1.1 EFFICIENT RECOGNITION OF AUTHENTIC DYNAMIC FACIAL EXPRESSIONS ON THE FEEDTUM DATABASE**

*Frank Wallhoff, Björn Schuller, Michael Hawellek, Gerhard Rigoll, Technische Universität München, Germany*

In order to allow for fast recognition of a user's effective state we discuss innovative holistic and self organizing approaches for efficient facial expression analysis. The feature set is thereby formed by global descriptors and MPEG based DCT coefficients. In view of subsequent classification we compare modelling by pseudo multidimensional Hidden Markov Models and Support Vector Machines. Within the latter case super-vectors are constructed based on Sequential Floating Search Methods. Extensive test-runs as a proof of concept are carried out on our publicly available FEEDTUM database consisting of elicited spontaneous emotions of 18 subjects within the MPEG-4 emotion-set plus added neutrality.

13:30 - 13:50

**MP1-L1.2 TV COMMERCIAL CLASSIFICATION BY USING MULTI-MODAL TEXTUAL INFORMATION**

*Yantao Zheng, National University of Singapore, Singapore; Lingyu Duan, Qi Tian, Institute for Infocomm Research, Singapore; Jesse S. Jin, University of Newcastle, Australia*

In this paper, we propose an approach for TV commercial video classification by the categories of advertised products or services (e.g. automobiles, healthcare products, etc). Since automatic speech recognition (ASR) and optical character recognition (OCR) can deliver meaningful textual information related to products or services, TV commercial video classification is formulated as the problem of text categorization. However, there exist two challenges. Firstly, the background music of TV commercials makes ASR techniques yield erroneous and deficient output transcripts. Secondly, even if ASR and OCR could work perfectly, the limited textual information from TV commercials do not suffice to train a generic and non-overfitting text categorizer. For the first issue, our approach resorts to the external resources to expand deficient ASR and OCR transcripts. The output transcripts of ASR and OCR are parsed to yield a few keywords, on which a Web searching is executed to retrieve relevant and semantically informative articles from World Wide Web (WWW). The retrieved articles are then utilized to construct textual feature vectors and perform text categorization on behalf of commercials. For the second issue, a topic-wise document corpus is constructed from the public corpora like Reuters-21578 or from the articles manually collected from WWW for the training of text categorizers. Experimental results have shown that the proposed approach alleviates the negative effects from weak ASR/OCR performance and yield a promising classification accuracy of 80.9%.

13:50 - 14:10

**MP1-L1.3 SUPPORT VECTOR MACHINE FOR MULTIPLE FEATURE CLASSIFICATION**

*Bing-Yu Sun, Moon-Chuen Lee, The Chinese University of Hong Kong, Hong Kong SAR of China*

This paper proposes an effective method for training a combined SVM classifier through combining several base classifiers, each for one feature, at the same time. Traditional combination methods normally train the base classifiers before combining them to form a combined classifier. While the proposed method needs to solve only one optimization problem formulated in this paper, the traditional methods need to solve several optimization problems, one for each base classifier for a single feature. Thus, the training complexity of the proposed method can be much reduced. Furthermore, the combination of the base classifiers in our proposed approach takes into account a base classifier's performance on the training data and the generalization ability of the combined classifier; whereas the traditional combination methods consider only a base classifier's performance on the training data. Besides the underlying lower training complexity, the experiment results show that the proposed combination method outperforms other approaches.

14:10 - 14:30

**MP1-L1.4 WEB IMAGE MINING BASED ON MODELING CONCEPT-SENSITIVE SALIENT REGIONS**

*Jing Liu, Qingshan Liu, Jinqiao Wang, Hanqing Lu, Songde Ma, Chinese Academy of Sciences, China*

In this paper, we propose a probabilistic model for web image mining, which is based on concept-sensitive salient regions without human intervene. Our goal is to achieve a middle-level understanding of image semantics to bridge the semantic gap existing in the field of image mining and retrieval. With the help of a popular search engine, semantically relevant images are collected, and concept-sensitive salient regions are extracted automatically based on an attention model. Then the semantic concept model is learned from the joint distribution of all salient regions with Gaussian Mixture Model and Expectation-Maximization algorithm. In addition, by incorporating semantically irrelevant un-salient regions as negative samples, the discriminative power of the solution is further enhanced. Experiments demonstrate the encouraging performance of the proposed method.

**MP1-L2 Perceptual Processing I (Special Session)**

Time: Monday, July 10, 13:10 - 14:30

Place: Johnston

Co-Chairs: Xiaokang Yang, Shanghai Jiao Tong University, Weisi Lin, Institute for Infocomm Research and Minoru Etoh, Multimedia Laboratories, NTT DoCoMo

13:10 - 13:30

**MP1-L2.1 INTERACTIONS AND INTEGRATIONS OF MULTIPLE SENSORY CHANNELS IN HUMAN BRAIN**

*Shin'ya Nishida, NTT Communication Science Laboratories, Japan*

Here I describe a couple of new principles with regard to interactions and integrations of multiple sensory channels in the human brain. First, as opposed to the general belief that the perception of shape and that of color are relatively independent of motion processing, human visual system integrates shape and color signals along perceived motion trajectory in order to improve visibility of shape and color of moving objects. Second, when the human sensory system binds the outputs of different sensory channels, (including audio-visual signals) based on their temporal synchrony, it uses only sparse salient features rather than using the time courses of full sensory signals. We believe these principles are potentially useful for development of effective audiovisual processing and presentation devices.

13:30 - 13:50

**MP1-L2.2 OPTIMIZING ALGORITHMS FOR REGION-OF-INTEREST VIDEO COMPRESSION, WITH APPLICATION TO MOBILE TELEHEALTH**

*Sira Rao, Nikil Jayant, Georgia Institute of Technology, United States*

Wireless communication of video poses constraints on information capacity. Region-of-Interest (ROI) video coding provides higher quality in the ROI, but poorer quality in the background (BKGRND), for a given total bitrate (TBR). Researchers, including the authors, have also proposed more graceful quality management methods, using what is referred to here as an Extended-Region-of-Interest (EROI). We consider three levels of losslessness – mathematical, diagnostic, and perceptual, with the goal of associating them with the above-mentioned regions. We describe work in progress aimed at optimizing an elastic expert system based on the above methodology, with telehealth video as its anchor. The optimizations are threefold – user, perceptual, and network oriented, and are incorporated in the rate control algorithm. We propose a rate control method where, unlike conventional methods, bit allocation is shifted from the frame level to individual regions within the frame. Thereafter, the above-mentioned criteria are used to determine regional bit allocation. Peak-Signal-to-Noise-Ratio (PSNR) results show, as expected, that the proposed scheme achieves higher ROI-EROI quality than the verification model VM8 of MPEG4. This is illustrated with four examples of pediatrics video. The value and design of the proposed methodology is being corroborated by subjective testing involving medical experts. We are independently researching another standing issue in the telehealth application, that of low complexity segmentation and tracking of the ROI-EROI boundaries.

13:50 - 14:10

**MP1-L2.3 A CAPTCHA BASED ON THE HUMAN VISUAL SYSTEMS MASKING CHARACTERISTICS**

*Rony Ferzli, Rida Bazzi, Lina J. Karam, Arizona State University, United States*

In this paper, a CAPTCHA is presented based on the masking characteristics of the Human Visual System (HVS). Knowing that noise can be masked by high activity regions and showing that edges can be masked by noise for a human observer while still being detected by machines, the suggested CAPTCHA is composed of English alphabets that are picked randomly and written with a combination of texture and edges with added noise such as to deceive the machine by randomly changing the visibility of characters for humans. The proposed CAPTCHA is highly legible and robust to brute-force attacks and sophisticated Object Character Recognition (OCR) segmentation algorithms.

14:10 - 14:30

#### **MP1-L2.4 PHOTOREALISTIC ATTENTION-BASED GAZE ANIMATION**

*Laurent Itti, Nitin Dhavale, Frédéric Pighin, University of Southern California, United States*

We apply a neurobiological model of visual attention and gaze control to the automatic animation of a photorealistic virtual human head. The attention model simulates biological visual processing along the occipito-parietal pathway of the primate brain. The gaze control model is derived from motion capture of human subjects, using high-speed video-based eye and head tracking apparatus. Given an arbitrary video clip, the model predicts visual locations most likely to attract an observer's attention, and simulates the dynamics of eye and head movements towards these locations. Tested on 85 video clips including synthetic stimuli, video games, TV news, sports, and outdoor scenes, the model demonstrates a strong ability at saccading towards and tracking salient targets. The resulting autonomous virtual human animation is of photorealistic quality.

### **MP1-L3: Monday, July 10, 13:10 - 14:30**

#### **MP1-L3 Content Understanding III (Lecture)**

Time: Monday, July 10, 13:10 - 14:30

Place: Thomson

Chair: Sridhar Krishnan, Ryerson University

13:10 - 13:30

#### **MP1-L3.1 IMAGE AUTO-ANNOTATION USING A STATISTICAL MODEL WITH SALIENT REGIONS**

*Jiayu Tang, Jonathon Hare, Paul Lewis, University of Southampton, United Kingdom*

Traditionally, statistical models for image auto-annotation have been coupled with image segmentation. Considering the performance of the current segmentation algorithms, it can be meaningful to avoid a segmentation stage. In this paper, we propose a new approach to image auto-annotation by building on previously developed statistical models. In this approach, segmentation is avoided through the use of salient regions. The use of the statistical model results in an annotation performance which improves upon our previously proposed saliency-based word propagation technique. We also show that the use of salient regions achieves better results than the use of general image regions or segments.

13:30 - 13:50

#### **MP1-L3.2 SAMPLING STRATEGIES FOR ACTIVE LEARNING IN PERSONAL PHOTO RETRIEVAL**

*Yi Wu, Igor Kozintsev, Jean-Yves Bouguet, Carole Dulong, Intel Corporation, United States*

With the advent and proliferation of digital cameras and computers, the number of digital photos created and stored by consumers has grown extremely large. This created increasing demand for image retrieval systems to ease interaction between consumers and personal media content. Active learning is a widely used user interaction model for retrieval systems, which learns the query concept by asking users to label a number of images at each iteration. In this paper, we study sampling strategies for active learning in personal photo retrieval. In order to reduce human annotation efforts in a content-based image retrieval setting, we propose using multiple sampling criteria for active learning: informativeness, diversity and representativeness. Our experimental results show that by combining multiple sampling criteria in active learning, the performance of personal photo retrieval system can be significantly improved.



13:50 - 14:10

**MP1-L3.3 A NEW STUDY ON DISTANCE METRICS AS SIMILARITY MEASUREMENT**

*Jie Yu, University of Texas, San Antonio, United States; Jaume Amores, IMEDIA Research Group, INRIA, France; Nicu Sebe, University of Amsterdam, Netherlands; Qi Tian, University of Texas, San Antonio, United States*

Distance metric is widely used in similarity estimation and the most popular Euclidean and Manhattan distance may not be suitable for all data distribution. In this paper, a general guideline to establish the relation between a distribution model and its corresponding similarity estimation is proposed. Based on Maximum Likelihood theory, we propose new distance metrics, such as harmonic distance and geometric distance. Because the feature elements maybe from heterogeneous sources and usually have different influence on similarity estimation, it is inappropriate to model the distribution as isotropic. We propose a novel boosted distance metric that not only finds the best distance metric that fits the distribution of the underlying elements but also selects the most important feature elements in respect to similarity. The boosted distance metric is tested on fifteen benchmark data sets from the UCI repository and two image retrieval applications. In all the experiments, robust results are obtained based on the proposed methods.

14:10 - 14:30

**MP1-L3.4 MATCHING FACES WITH TEXTUAL CUES IN SOCCER VIDEOS**

*Marco Bertini, Alberto Del Bimbo, Walter Nunziati, Università di Firenze, Italy*

In soccer videos, most significant actions are usually followed by close-up shots of players that take part in the action itself. Automatically annotating the identity of the players present in these shots would be considerably valuable for indexing and retrieval applications. Due to high variations in pose and illumination across shots however, current face recognition methods are not suitable for this task. We show how the inherent multiple media structure of soccer videos can be exploited to understand the players' identity without relying on direct face recognition. The proposed method is based on a combination of interest point detector to "read" textual cues that allow to label a player with its name, such as the number depicted on its jersey, or the superimposed text caption showing its name. Players not identified by this process are then assigned to one of the labeled faces by means of a face similarity measure, again based on the appearance of local salient patches. We present results obtained from soccer videos taken from various recent games between national teams.

**MP1-L4: Monday, July 10, 13:10 - 14:30**

**MP1-L4 Authentication and Forensic Analysis (Lecture)**

Time: Monday, July 10, 13:10 - 14:30

Place: Toronto I

Chair: Shih-Fu Chang, Columbia University

13:10 - 13:30

**MP1-L4.1 A CONTENT-AWARE STREAM AUTHENTICATION SCHEME OPTIMIZED FOR DISTORTION AND OVERHEAD**

*Zhishou Zhang, Qibin Sun, Wai-Choong Wong, Institute for Infocomm Research, Singapore; John Apostolopoulos, Susie Wee, Hewlett-Packard Laboratories, United States*

This paper proposes a content-aware authentication scheme optimized to account for distortion and overhead for media streaming. When authenticated media is streamed over a lossy network, a received packet is consumed only when it is both decodable and authenticated. In most media formats, some packets are more important than others. This naturally motivates allocating more redundant authentication information for the more important packets in order to maximize their probability of authentication and thereby minimize distortion at the receiver. Toward this goal, with awareness of the media content, we formulate an optimization framework to compute an authentication graph to maximize the expected media quality at the receiver, given specific authentication overhead and knowledge of network loss rates. Experimental results with JPEG-2000 coded images demonstrate that the proposed method achieves our design goal in that the R-D curve of the authenticated image is very close to the R-D curve when no authentication is required.



13:30 - 13:50

**MP1-L4.2 AUTHENTICATING MULTIMEDIA TRANSMITTED OVER WIRELESS NETWORKS: A CONTENT-AWARE STREAM-LEVEL APPROACH**

*Zhi Li, Yong Lian, National University of Singapore, Singapore; Qibin Sun, Institute for Infocomm Research, Singapore; Chang Wen Chen, Florida Institute of Technology, United States*

We propose in this paper a novel content-aware stream-level approach to authenticating multimedia data transmitted over wireless networks. The proposed approach is fundamentally different from conventional authentication methods and offers robust authentication for multimedia data in the presence of channel noise. The scheme is designed in such a way that it facilitates explicit capture and exploitation of channel condition as well as how the multimedia content is packetized and transmitted. The design allows the integration of authentication with the framework of Joint Source and Channel Coding (JSCC) to achieve adaptiveness to the content and efficient utilization of limited bandwidth. We have realized the proposed scheme through optimal resource allocation and authentication graph construction. Experiment results demonstrated the effectiveness of this novel approach.

13:50 - 14:10

**MP1-L4.3 DETECTING IMAGE SPLICING USING GEOMETRY INVARIANTS AND CAMERA CHARACTERISTICS CONSISTENCY**

*Yu-Feng Hsu, Shih-Fu Chang, Columbia University, United States*

Recent advances in computer technology have made digital image tampering more and more common. In this paper, we propose an authentic vs. spliced image classification method making use of geometry invariants in a semi-automatic manner. For a given image, we identify suspicious splicing areas, compute the geometry invariants from the pixels within each region, and then estimate the camera response function (CRF) from these geometry invariants. The cross-fitting errors are fed into a statistical classifier. Experiments show a very promising accuracy, 87%, over a large data set of 363 natural and spliced images. To the best of our knowledge, this is the first work detecting image splicing by verifying camera characteristic consistency from a single-channel image.

14:10 - 14:30

**MP1-L4.4 VIDEO BASED PERSON AUTHENTICATION VIA AUDIO/VISUAL ASSOCIATION**

*Ming Liu, Thomas S. Huang, University of Illinois at Urbana-Champaign, United States*

Multi-modal person authentication systems can achieve higher performance and robustness by combining different modalities. The current fusion strategies of different modalities are mainly based on the output of individual modalities. However, there are detail structures between facial movement and speech signal. In this paper, Audio/Visual association, a lower level fusion, is proposed to fuse the information between lip movement and speech signal. The experimental results indicate that this type of fusion strategy improve the performance of multi-modal person authentication system.

**MP1-L5: Monday, July 10, 13:10 - 14:30**

**MP1-L5 P2P Multimedia I (Special Session)**

Time: Monday, July 10, 13:10 - 14:30

Place: Toronto II

Co-Chairs: Zhu Liu, AT&T Research Labs and Deepa Kundur, Texas A&M University

13:10 - 13:30

**MP1-L5.1 ON PEER-TO-PEER MULTIMEDIA CONTENT ACCESS AND DISTRIBUTION**

*Zhu Liu, AT&T Research Labs, United States; Heather Yu, Panasonic Princeton Research Lab, United States; Deepa Kundur, Texas A&M University, United States; Madjid Merabti, Liverpool John Moores University, United Kingdom*

This paper provides a brief overview of recent progress of peer-to-peer (P2P) technologies for multimedia applications. We provide an overview of the technical challenges, creative solutions and results related to P2P file sharing and content search, and P2P media streaming.

13:30 - 13:50

**MP1-L5.2 ADAPTIVE ERASURE RESILIENT CODING IN DISTRIBUTED STORAGE**

*Jin Li, Microsoft Research, United States*

A challenge of peer-to-peer (P2P) storage is to build reliable data storage from inherently unreliable P2P network, and to do so efficiently. In this paper, we investigate the use of adaptive erasure resilient code (ERC) which changes ERC fragment size according to the size of the file distributed. We show that adaptive ERC may greatly improve the efficiency and reliability of P2P storage. A number of design policies and strategies for P2P storage application are also investigated.

13:50 - 14:10

**MP1-L5.3 EFFICIENT SEARCH IN P2P-BASED VIDEO-ON-DEMAND STREAMING SERVICE**

*HuiCheng Chi, Qian Zhang, Hong Kong University of Science and Technology, Hong Kong SAR of China*

Providing video-on-demand streaming service to a large population of clients using peer-to-peer approach is drawing great interest recently. Since clients' demands are asynchronous and the buffered contents are continuously changing, how to find partners with expected data and collaborate with each other for future content delivery are very important and challenging. In this paper, we propose a generic buffer-assisted search (BAS) scheme to improve partner search efficiency. Extensive simulation results demonstrate that BAS can provide faster response time with less control cost than the existing search methods.

14:10 - 14:30

**MP1-L5.4 LOW LATENCY VIDEO STREAMING OVER PEER-TO-PEER NETWORKS**

*Eric Setton, Jeonghun Noh, Bernd Girod, Stanford University, United States*

We study peer-to-peer multicast streaming, where a source distributes real-time video to a large population of hosts by making use of their forwarding capacity rather than relying on dedicated media servers. We present a distributed streaming protocol which builds and maintains multiple multicast trees. The protocol is combined with an adaptive scheduling algorithm which ensures packets destined to a large number of peers, or particularly important to decode the video, are sent in priority. Experiments carried out over a simulated network of up to 3000 peers illustrate the performance of the protocol. For low latency video streaming, the prioritization algorithm offers large performance gains, especially for large audiences.

**MP1-L6: Monday, July 10, 13:10 - 14:30**

**MP1-L6 Multimedia Computing and Software I (Lecture)**

Time: Monday, July 10, 13:10 - 14:30

Place: Varley

Chair: Jean-Philippe Farrugia, University of Lyon

13:10 - 13:30

**MP1-L6.1 MODELING ADAPTIVE MEDIA PROCESSING WORKFLOWS**

*K. Selçuk Candan, Gisik Kwon, Lina Peng, Arizona State University, United States; Maria Luisa Sapino, Università di Torino, Italy*

ARIA, ARchitecture for Interactive Arts, is a middleware to process, filter, and fuse sensory inputs and actuate responses in real-time. An ARIA media workflow graph describes how the data sensed through media will be processed and what audiovisual responses will be actuated. Each object streamed between ARIA processing components is subject to transformations, as described by a media workflow graph. The media capture and processing components, such as media filters and fusion operators, are programmable and adaptable; i.e, the delay, size, frequency, and quality/precision characteristics of individual operators can be controlled via a number of parameters. In this paper, we present the underlying model which captures the dynamic nature of the ARIA media processing workflows. This model enables design time verification, optimization, and runtime adaptation.

13:30 - 13:50

**MP1-L6.2 MODEL AND USAGE OF A CORE MODULE FOR AXMEDIS/MPEG-21 CONTENT MANIPULATION TOOLS**

*Pierfrancesco Bellini, Paolo Nesi, L. Ortimini, Davide Rogai, Andrea Vallotti, University of Florence, Italy*

Despite the relevance of the MPEG-21 standard, little has been done about the modeling of authoring tools and players for the production/consumption of MPEG-21 digital objects. The design and implementation of MPEG-21 tools present several critical points to be solved at the architectural level, so as to ensure security and provide a suitable support and the requested flexibility for manipulating any kind of digital resource, according to the spirit of MPEG-21 standard. This paper presents the AXMEDIS core model and components for MPEG-21 content authoring tools and players. The proposed architecture provides both data manipulation flexibility and a high level of security when digital resources are used. The same architecture can be used to develop any other MPEG-21 DI tools. The architecture presented has been adopted to build authoring and player tools developed for the AXMEDIS IST FP6 R&D European Commission project (<http://www.axmedis.org>).

13:50 - 14:10

**MP1-L6.3 HIERARCHICAL LOAD TESTING ARCHITECTURE USING LARGE SCALE VIRTUAL CLIENTS**

*Bum Hyun Lim, Jin Ryong Kim, Kwang Hyun Shim, Electronics and Telecommunications Research Institute (ETRI), Republic of Korea*

In this work, we develop a hierarchical load testing architecture using large scale virtual clients to reduce the testing time and ensure the stability of the server for distributed applications. It explicitly secures the stability of the servers for networked virtual environment and at the same time, it elaborately generates actual loads for testing the performance of the servers. Our agent based load testing architecture provides variety of interactions of virtual entities in the virtual worlds to perform realistic simulations. Simulation results illustrate that our proposed architecture ensures the stability and capacity of the servers for both massively multiplayer online games and peer-to-peer network games.

14:10 - 14:30

**MP1-L6.4 GPUCV: A FRAMEWORK FOR IMAGE PROCESSING ACCELERATION WITH GRAPHICS PROCESSORS**

*Jean-Philippe Farrugia, LIRIS - Université Lyon 1, France; Patrick Horain, Erwan Guehenneux, Yannick Alusse, Institut National des Télécommunications, France*

This paper presents a state of the art report on using graphics hardware for image processing and computer vision. Then we describe GPUCV, an open library for easily developing GPU accelerated image processing and analysis operators and applications.

**MP1-P1: Monday, July 10, 13:30 - 14:30**

**MP1-P1 Wireless and Mobile Multimedia (Poster)**

Time: Monday, July 10, 13:30 - 14:30

Place: Toronto III

Chair: Nikolaos Boulgouris, King's College London

**MP1-P1.1 AN IMPROVED RESOURCE RESERVATION ALGORITHM FOR IEEE 802.15.3**

*Xin Liu, Qionghai Dai, Qiufeng Wu, Tsinghua University, China*

In this paper, we propose an improved resource reservation algorithm ESRPT (Enhanced Shortest Remaining Processing Time) for bursty traffic based on IEEE 802.15.3 standard. In this algorithm, each transmitter reports current fragments number of the first MSDU (MAC Service Data Unit) and all of the fragments number of the remainder MSDUs in the pending transmission queue to PNC (piconet coordinator). In the next superframe, PNC firstly allocates part CTAs (Channel Time Allocation) for each stream based on the remainder fragments number of the first MSDU by SRPT rule, then allocates remainder CTAs for each stream based on all fragments number of remainder MSDUs by the same SRPT rule. This algorithm decreases the job failure rates of delay-sensitive video streaming efficiently at the cost of slightly increasing communication overheads. Simulation results show that our proposed ESRPT method achieves better performance in QoS for multimedia streams compared to the existing SRPT schemes.

**MP1-P1.2 ROBUST TRANSMISSION OF H.264/AVC VIDEO USING ADAPTIVE SLICE GROUPING AND UNEQUAL ERROR PROTECTION**

*Nikolaos Thomos, Savvas Argyropoulos, Informatics and Telematics Institute, Greece; Nikolaos V. Boulgouris, King's College London, United Kingdom; Michael G. Strintzis, Informatics and Telematics Institute, Greece*

We present a novel scheme for the transmission of H.264/AVC video streams over lossy packet networks. The proposed scheme exploits the error resilient features of H.264/AVC codec and employs Reed-Solomon codes to protect effectively the streams. The optimal classification of macroblocks into slice groups and the optimal channel rate allocation are achieved by iterating two interdependent steps. Simulations clearly demonstrate the superiority of the proposed method over other recent algorithms for transmission of H.264/AVC streams.

**MP1-P1.3 REUSE OF MOTION PROCESSING FOR CAMERA STABILIZATION AND VIDEO CODING**

*Bao Lei, Rene Klein Gunnewiek, Philips Research Laboratories, Netherlands; Peter H. N. de With, University of Technology Eindhoven, Netherlands*

The low bit rate of existing video encoders relies heavily on the accuracy of estimating actual motion in the input video sequence. In this paper, we propose a Video Stabilization and Encoding (ViSE) system to achieve a higher coding efficiency through a preceding motion processing stage (to the compression), of which the stabilization part should compensate for vibrating camera motion. The improved motion prediction is obtained by differentiating between the temporal coherent motion and a more noisy motion component which is orthogonal to the coherent one. The system compensates the latter undesirable motion, so that it is eliminated prior to video encoding. To reduce the computational complexity of integrating a digital stabilization algorithm with video encoding, we propose a system that reuses the already evaluated motion vector from the stabilization stage in the compression. As compared to H.264, our system shows a 14% reduction in bit rate yet obtaining an increase of about 0.5 dB in SNR.

**MP1-P1.4 VIDEO ENCODING AND SPLICING FOR TUNE-IN TIME REDUCTION IN IP DATACASTING (IPDC) OVER DVB-H**

*Mehdi Rezaei, Tampere University of Technology, Finland; Miska M. Hannuksela, Nokia Research Center, Finland; Moncef Gabbouj, Tampere University of Technology, Finland*

A novel video encoding and splicing method is proposed which minimizes the tune-in time of “channel zapping”, i.e. changing from one audiovisual service to another, in IPDC over Digital Video Broadcasting for Handheld terminals (DVB-H). DVB-H uses a time-sliced transmission scheme to reduce the power consumption used for radio reception. Tune-in time in DVB-H refers to the time between the start of the reception of a broadcast signal and the start of the media rendering. One of the significant factors in tune-in time is the time from the start of media decoding to the start of correct output from decoding, which is minimized when a time-slice is started with a random access point picture such as an independent decoding refresh (IDR) picture in H.264/AVC. In IPDC over DVB-H, encapsulation to time-slices is performed independently from encoding in a network element called IP encapsulator. At the time of encoding, time-slice boundaries are not known exactly, and it is therefore impossible to govern the location of IDR pictures relative to time-slices. It is proposed that an additional stream consisting of IDR pictures only is transmitted to the IP encapsulator, which replaces pictures in a normal bitstream with IDR pictures according to time-slice boundaries in order to achieve the minimum tune-in time. It has to be ensured that the “spliced” stream resulting from the operation of the IP encapsulator complies with the Hypothetical Reference Decoder (HRD) specification of H.264/AVC. A video encoding and rate control system is proposed to satisfy the HRD requirements for the spliced stream. Simulation results show that in addition to fulfilling HRD compliancy, good average quality of decoded video is achieved with minimum tune-in time.

**MP1-P1.5 EFFICIENT ALGORITHM FOR GLOBALLY OPTIMAL UNEVEN ERASURE-PROTECTED PACKETIZATION OF SCALABLE CODESTREAMS**

*Sorina Dumitrescu, Xiaolin Wu, Zhe Wang, McMaster University, Canada*

A new algorithm is presented for rate-fidelity optimal packetization of scalable source bit streams with uneven erasure protection. It provides the globally optimal solution for input sources of convex rate-fidelity function and for a wide class of erasure channels, including channels for which the probability of losing  $n$  packets is monotonically decreasing in  $n$ , and independent erasure channels with packet erasure rate smaller than 0.5. The time and space complexities of the new algorithm are both  $O(NL)$ , where  $N$  is the number of packets and  $L$  is the packet payload size, comparing to the  $O(NL^2)$  time and space complexities of the existing globally optimal solution. When applied to SPIHT compressed images, the results of the proposed algorithm are virtually the same as the globally optimal.

**MP1-P1.6    MODELING AND PERFORMANCE ANALYSIS OF INITIAL CONNECTION IN IEEE 802.16 PMP NETWORKS**

*Lidong Lin, Bo Han, Weijia Jia, City University of Hong Kong, Hong Kong SAR of China*

In this paper, we propose an accurate analytical model to analyze the performance of initial ranging requests in IEEE 802.16 networks. Two metrics, connection probability and average connection delay, are investigated to evaluate the network performance. Performance observation demonstrate that the connection probability is not heavily influenced by the contention window size and reconnection retry limitation but the average connection delay is sensitive to the above two parameters. Moreover, we find that improving the service capacity and buffer size of base station can optimize the connection probability and average connection delay.

**MP1-P1.7    UNEQUAL ITERATIVE DECODING FOR EFFICIENT VIDEO TRANSMISSION**

*Yongfang Wang, Songyu Yu, Xiaokang Yang, Shanghai Jiao Tong University, China*

We present an Unequal Iterative Decoding (UID) approach for minimization of the receiver power consumption subject to a given quality of service, by exploiting data partitioning and turbo decoding. We assign unequal decoding iterations of forward error correction (FEC) to data partitions with different importance by jointly considering the source coding, channel coding and receiver power allocation. The proposed scheme has been applied to H.264 video over AWGN channel, and achieves excellent tradeoff between video delivery quality and power consumption, and yields significant power saving compared with the classical equal iterative decoding (EID) approach in wireless video transmission.

**MP1-P1.8    A PRIORITY-BASED ADAPTIVE SCHEME FOR WIRELESS MULTIMEDIA DELIVERY**

*Edward Casey, Gabriel-Miro Muntean, Dublin City University, Ireland*

In wireless multimedia streaming, there is a need to allow for client prioritisation in order to enable provision of end user perceived quality in direct relationship with client device importance. Currently, the same priority is given to all clients, independent of their characteristics, often resulting in unfair distribution of throughput. This paper proposes a priority-based wireless adaptive multimedia delivery scheme that enables client prioritisation during multimedia distribution over IP networks. The paper presents simulation results outlining the benefits of applying the algorithm, illustrating the improvement in bandwidth allocation and in overall end user perceived quality. The algorithm focuses on a residential wireless local area network and assigns static priorities based on device characteristics.

**MP1-P1.9    SEAMLESS MULTIMEDIA TRANSMISSION DURING FAST HANDOVER IN MOBILE IPV6 ENVIRONMENT**

*Jin Ryong Kim, Electronics and Telecommunications Research Institute (ETRI), Republic of Korea; Youn-Hee Han, Korea University of Technology and Education, Republic of Korea; Kwang Hyun Shim, Electronics and Telecommunications Research Institute (ETRI), Republic of Korea*

In this paper, we propose the seamless video transmission scheme over wireless LANs during fast handover in Mobile IPv6. In fact, a mobile node cannot receive the IP packets during handover. A number of extensions to fast handover for mobile IPv6 have been introduced to recover its weakness. Fast handover algorithm in Mobile IPv6 allows the new access router in the adjacent cell to pre-buffer the data packets of a mobile node to prevent packet loss. In a mobile communication environment, however, transmission of multimedia data should consider the realtimeliness of the data. We focus on the multimedia data delivery during fast handover in Mobile IPv6 environment.

**MP1-P1.10    ADVANCED WIRELESS MULTIUSER VIDEO STREAMING USING THE SCALABLE VIDEO CODING EXTENSIONS OF H.264/MPEG-4 AVC**

*Günther Liebl, Technical University of Munich, Germany; Thomas Schierl, Thomas Wiegand, Fraunhofer – Heinrich Hertz Institute, Germany; Thomas Stockhammer, Nomor Research, Germany*

We propose a dynamic sharing of radio resources in a wireless multiuser system by combining a flexible video coding scheme, namely the currently developed scalable video coding (SVC) extension of H.264/AVC with appropriate radio link buffer management for multiuser streaming services. The paper introduces the components of the proposed system and the used cross-layer interface. Simulation results verify the benefits of our approach.

### **MP1-P1.11 A HANDOVER APPROACH TO DVB-H SERVICES**

*Ville Ollikainen, Chengyuan Peng, VTT, Finland*

DVB-H offers a new platform for IP-based services and contributes to universal access. There are many challenges of providing multimedia TV experience for DVB-H users anywhere anytime. One big challenge is due to the strict reception conditions and mobility. For example, when a mobile receiver enters a new cell with different transmission frequency, the receiver must accomplish a seamless handover process in order to continue the selected service without an interrupt. Furthermore, it is possible to build DVB-H services into existing DVB-T networks that provide even less possibilities to optimize the transmission for mobile use. This paper presents a novel handover approach to IP streams on a DVB-H network. We handover IP streams by switching between the DVB-H and the UMTS networks without doing any frequency scan. We have successfully tested the service handover approach. The validation has shown that there was no any packet loss during a handover process. The method also provides technology about efficient sharing of wireless network infrastructure with DVB-H, and makes it more feasible to include DVB-H services into existing DVB-T networks and can be used for reducing the time for changing a channel (zapping time).

### **MP1-P1.12 ROBUST VIDEO TRANSMISSION OVER MIMO-OFDM SYSTEM USING MDC AND SPACE TIME CODES**

*Haifeng Zheng, Fuzhou University, China; Congchong Ru, Tsinghua University, China; Lun Yu, Fuzhou University, China; Chang Wen Chen, Florida Institute of Technology, United States*

MIMO-OFDM is a promising technique for the broadband wireless communication system. In this paper, we propose a novel scheme that integrates multiple description coding (MDC), error resilient video coding, and unequal error protection strategy with various space time coding codes for robust video transmission over MIMO-OFDM system. The proposed MDC coder generates multiple bitstreams of equal importance which are very suitable for multiple antennas system. Furthermore, according to the contribution to the reconstructed video quality, we apply unequal error protection strategy using BLAST and STBC space time codes for each video bitstream. Experimental results have demonstrated that the proposed scheme can achieve desired trade-off between the reconstructed video quality and the transmission efficiency.

## **MP1-P2: Monday, July 10, 13:30 - 14:30**

### **MP1-P2 Multimedia Applications I (Poster)**

Time: Monday, July 10, 13:30 - 14:30

Place: Toronto III

Chair: Mihaela van der Schaar, University of California, Los Angeles

### **MP1-P2.13 A THREE-LAYER VIRTUAL DIRECTOR MODEL FOR SUPPORTING AUTOMATED MULTI-SITE DISTRIBUTED EDUCATION**

*Bin Yu, University of Illinois at Urbana-Champaign, United States; Cha Zhang, Yong Rui, Microsoft Research, United States; Klara Nahrstedt, University of Illinois at Urbana-Champaign, United States*

In multi-site distributed education (MSDE), video streams from multiple sites are available. To best utilize the limited screen space at each site, we develop a customizable, automated display management system in this paper, i.e., only user-preferred streams will be shown as triggered by events and timers. The configuration of such user preference, however, is challenging because it has to be both human-friendly and machine-friendly. To address this challenge, we propose a three-layer virtual director model. In the user layer, we identify three categories of parameters that can represent a wide range of user preferences yet are easy to use. These preferences are then automatically translated into a machine-friendly timed automaton in the execution layer. The automaton is simulated dynamically, which selects a subset of streams to show on the screen through a display layer. Evaluation results demonstrate the correctness and efficiency of the proposed framework.



#### **MP1-P2.14 GENERATING A TIME SHRUNK LECTURE VIDEO BY EVENT DETECTION**

*Takao Yokoi, Hironobu Fujiyoshi, Chubu University, Japan*

Streaming a lecture video via the Internet is important for Elearning. We have developed a system that generates a lecture video using virtual camerawork based on shooting techniques of broadcast cameramen. However, viewing a fulllength video takes time for students. In this paper, we propose a method for generating a time shrunk lecture video using event detection. We detect two kinds of events: a speech period and a chalkboard writing period. A speech period is detected by voice activity detection with LPC cepstrum and classified into speech or non-speech using Mahalanobis distance. To detect chalkboard writing periods, we use a graph cuts technique to segment a precise region of interests such as an instructor. By deleting content-free periods, i.e, period without the events of speech and writing, and fast-forwarding writing periods, our method can generate a time shrunk lecture video automatically. The resulting generated video is about 20%~30% shorter than the original video in time. This is almost the same as the results of manual editing by a human operator.

#### **MP1-P2.15 GPU ACCELERATED INVERSE PHOTON MAPPING FOR REAL-TIME SURFACE REFLECTANCE MODELING**

*Takashi Machida, Osaka University, Japan; Naokazu Yokoya, Nara Institute of Science and Technology, Japan; Haruo Takemura, Osaka University, Japan*

This paper investigates the problem of object surface reflectance modeling, which is sometimes referred to as {it inverse reflectometry}, for photorealistic rendering and effective multimedia applications. A number of methods have been developed for estimating object surface reflectance properties in order to render real objects under arbitrary illumination conditions. However, it is still difficult to densely estimate surface reflectance properties in real-time. This paper describes a new method for real-time estimation of the non-uniform surface reflectance properties in the inverse rendering framework. Experiments are conducted in order to demonstrate the usefulness and the advantage of the proposed methods through comparative study.

#### **MP1-P2.16 MULTIMEDIA INFORMAL COMMUNICATION BY WEARABLE COMPUTER BASED ON REAL-WORLD CONTEXT AND GRAFFITI**

*Shun'ichi Tano, Tatsuhiro Takayama, Mitsuru Iwata, Tomonori Hashiyama, University of Electro-Communications, Japan*

The informal information can be so intuitive, emotional, personal, honest and subjective that it is very difficult to get them from the formal media. In this paper, we demonstrate a multimedia informal communication support system whose basic architecture is a combination of a wearable computer, a see-through head mounted display, a bone conduction speaker, and a 3D-pen. The user can create a "multimedia" graffiti such as a handwritten memo or a voice memo and place it in a real-world "multimedia" context such as at a certain location (by GPS), environmental sound, the real object (by RFID) and so on. The multimedia informal information is thus shared with others on the peer-to-peer self-organized network. We demonstrated the prototype system and conducted a preliminary evaluation.

#### **MP1-P2.17 PREDICTION-BASED GESTURE DETECTION IN LECTURE VIDEOS BY COMBINING VISUAL, SPEECH AND ELECTRONIC SLIDES**

*Feng Wang, Hong Kong University of Science and Technology, Hong Kong SAR of China; Chong-Wah Ngo, City University of Hong Kong, Hong Kong SAR of China; Ting-Chuen Pong, Hong Kong University of Science and Technology, Hong Kong SAR of China*

This paper presents an efficient algorithm for gesture detection in lecture videos by combining visual, speech and electronic slides. Besides accuracy, response time is also considered to cope with the efficiency requirements of real-time applications. Candidate gestures are first detected by visual cue. Then we modify HMM models for complete gestures to predict and recognize incomplete gestures before the whole gestures paths are observed. Gesture recognition is used to verify the results of gesture detection. The relations between visual, speech and slides are analyzed. The correspondence between speech and gesture is employed to improve the accuracy and the responsiveness of gesture detection.

#### **MP1-P2.18 THE INTERACTIVE COOKING SUPPORT SYSTEM IN MIXED REALITY ENVIRONMENT**

*Arata Horie, Satoru Mega, Kuniaki Uehara, Kobe University, Japan*

Recently, many learning systems, such as e-learning and WBT (Web Based Teaching) systems have been developed. In these systems, users can get educational contents and graphical material in a remote place through the network. However, the learning is always one-sided. That is, all users learn the same contents and cannot convey their states like "What is he or she doing now?" to these systems. Therefore, these systems are not necessarily satisfying all the users' demands. In this paper, we propose an interactive learning application of cooking in mixed reality (MR) environment. In an MR environment, a system should not only provide the useful knowledge, but also recognize users' state. Furthermore, by making a parallel transition model about cooking, it is possible to control the user's cooking process. Therefore, users can learn details about cooking more flexibly and effectively.

### **MP1-P2.19 INQUIRING OF THE SIGHTS FROM THE WEB VIA CAMERA MOBILES**

*Yinghua Zhou, Xin Fan, University of Science and Technology of China, China; Xing Xie, Microsoft Research Asia, China; Yuchang Gong, University of Science and Technology of China, China; Wei-Ying Ma, Microsoft Research Asia, China*

In this paper, we presented an image search service for mobile users. It can be used to acquire related information by taking and sending pictures to the server, for example, getting book reviews by a photo of the cover. The key problem here is to find images that contain the same prominent object as that in the query image. In the literature, local feature based image matching has been proven to outperform those based on global features. When using local features, however, one query image may contain thousands of high dimensional feature vectors. Each feature vector needs to match against millions of features in the database. Therefore, it is critical to design an efficient search scheme. Our proposed matching approach was based on identifying semi-local visual parts from multiple query images. Experiments on two real-world datasets showed that this approach was superior to conventional solutions.

### **MP1-P2.20 VIRTUAL DENTAL PATIENT: A SYSTEM FOR VIRTUAL TEETH DRILLING**

*Ioannis Marras, Leontios Papaleontiou, Nikolaos Nikolaidis, Kleoniki Lyroutdia, Ioannis Pitas, Aristotle University of Thessaloniki, Greece*

This paper introduces, a virtual teeth drilling system named Virtual Dental Patient designed to aid dentists in getting acquainted with the teeth anatomy, the handling of drilling instruments and the challenges associated with the drilling procedure. The basic aim of the system is to be used for the training of dental students. The application features a 3D model of the face and the oral cavity that can be adapted to the characteristics of a specific person and animated. Drilling using a haptic device is performed on realistic teeth models (constructed from real data), within the oral cavity. Results and intermediate steps of the drilling procedure can be saved for future use.

### **MP1-P2.21 PROVISIONING CONTEXT-AWARE ADVERTISEMENTS TO WIRELESS MOBILE USERS**

*Qusay Mahmoud, University of Guelph, Canada*

Mobile advertising, which is an area of mobile commerce, is a form of advertising that targets users of handheld wireless devices such as mobile phones and personal digital assistants (PDAs). Given the constraints of such devices, mobile advertisements should be context-aware in the sense that the behaviour of such value-added services is mostly driven by information based on user's location, time, user preferences, and the task at hand. This paper presents the design and prototyping of a system for provisioning context-aware advertisements to wireless mobile users. We use mobile agents because such autonomous entities have characteristics that can benefit mobile devices and the wireless environment. We have constructed a proof of concept implementation using Java technologies with support for WAP-enabled and J2ME-enabled devices.

## **MP1-P3: Monday, July 10, 13:30 - 14:30**

### **MP1-P3 H.264 II (Poster)**

Time: Monday, July 10, 13:30 - 14:30

Place: Toronto III

Chair: Panos Nasiopoulos, University of British Columbia

### **MP1-P3.22 A NOVEL TWO-STAGE RATE CONTROL SCHEME FOR H.264**

*Do-Kyoung Kwon, Mei-Yin Shen, C.-C. Jay Kuo, University of Southern California, United States*

A new rate control algorithm for the H.264 encoder is developed in this work. There are several unique features in the proposed rate control scheme. First, a two-stage encoding mechanism is used to resolve the problem of inter-dependency between R-D optimization (RDO) and rate control (RC) in H.264. Second, it is emphasized that the bits of the header information may occupy a larger portion of the total bit budget especially when encoding at low bit rates in H.264. To address this issue, a rate model for the header information is established so as to estimate header bits more accurately. Third, a new source rate model is proposed to estimate the quantization parameter (QP) from the residual signal. Built upon the above ideas, a rate control algorithm is developed for the H.264 baseline-profile encoder under the constant-bit-rate (CBR) constraint. It is shown by experimental results that the proposed scheme can control bit rates accurately with the R-D performance better than that of the rate control algorithm implemented in H.264 JM8.1a.



### **MP1-P3.23 PERCEPTUAL QUALITY METRIC FOR H.264 LOW BIT RATE VIDEOS**

*Ee Ping Ong, Weisi Lin, Zhongkang Lu, Susu Yao, Mei Hwan Loke, Institute for Infocomm Research, Singapore*

This paper proposed an objective video quality metric designed for automatically assessing the perceived quality of digitally compressed multimedia videos using H.264 video compression. The rationale in proposing perceptual-based metric is because traditional measure, peak signal-to-noise ratio (PSNR), has been found to correlate poorly with subjective quality ratings, particularly at much lower bit rates. In this paper, computational models have been applied to emulate human visual perception based on a combination of local and global modulating factors. The proposed video quality metric has been tested on CIF and QCIF video sequences compressed using H.264 video compression technique at various bit rates (24-384 Kbps) and frame rates (7.5-30Hz). Performance of the proposed metric with respect to subjective scores will be reported and a comparison with PSNR and also the video structural similarity method (being one of the best video quality metric for high bit rate videos recently reported in the literature) will also be provided in this paper.

### **MP1-P3.24 IMPROVEMENTS ON RATE-DISTORTION PERFORMANCE OF H.264 RATE CONTROL IN LOW BIT RATE VIDEO CODING**

*ShuMin Zhou, JinTao Li, YongDong Zhang, Chinese Academy of Sciences, China*

This paper points out some defects in the techniques used in H.264 rate control and presents two new techniques to improve them. The improved scheme has the following main features: 1) the bits allocated to each P-frame is proportional to the local motion in it, i.e., more bits are allocated to a frame if the local motion in it is stronger; 2) the quantization parameter (QP) calculation is based on a simple encoding complexity prediction scheme, which is more robust and simple than the quadratic model used by H.264 in low bit rate video coding. Simulation results show that compared to rate control scheme in H.264, the improved scheme has significantly improved R-D performance (up to 1.29dB).

### **MP1-P3.25 OFF-LINE MOTION DESCRIPTION FOR FAST VIDEO STREAM GENERATION IN MPEG-4 AVC/H.264**

*Yi Wang, University of Science and Technology of China, China; Xiaoyan Sun, Feng Wu, Shipeng Li, Microsoft Research Asia, China; Houqiang Li, Zhengkai Liu, University of Science and Technology of China, China*

The rate-distortion optimal mode decision as well as motion estimation adopted in H.264 brings a big challenge to real-time encoding and transcoding duo to the high computation complexity. In this paper, we propose a hierarchical motion description model to present the motion data of each macroblock (MB) from coarsely to finely. A preprocessing approach is developed to estimate the motion data for each MB at each quality level with regard to its reference quality, its adjacent MBs and the target bit-rate. The resulting motion data can be coded and stored as metadata in a media file or a stream. Moreover, we propose a method to readily extract the specific motion data from the model for each MB at given bit-rates. Experimental results have shown the effectiveness of our proposed motion description model in terms of coding efficiency as well as fast bit-rate adaptation in comparison with that of H.264.

### **MP1-P3.26 A CONTEXT-BASED ERROR DETECTION STRATEGY INTO H.264/AVC CABAC**

*Yang Li, Hongkai Xiong, Li Song, Songyu Yu, Shanghai Jiao Tong University, China*

Various error control schemes have been addressed in wireless video stream transmission. By combining an adaptive binary arithmetic coding technique with context modeling, CABAC as a normative part of H.264/AVC has achieved a high degree of adaptation and redundancy reduction. However, error propagation still remains a problem because of the property of arithmetic coding. The presented scheme compares the various error detection methods, and proposes an efficient error detection technique based on CABAC semantics, which is achieved by inserting detective markers denoting by syntax elements. The misdetection probability versus stream size expansion can be easily handled. In addition, placements of markers can vary with regard to specific video content, thus efficiency within this scheme is enhanced. Comparison with other detection scheme is also presented.

### **MP1-P3.27 JOINT ADAPTIVE INTRA REFRESHMENT AND UNEQUALLY ERROR PROTECTION ALGORITHMS FOR ROBUST TRANSMISSION OF H.264/AVC VIDEO**

*Jin Xu, Zhimei Wu, Chinese Academy of Sciences, China*

In this paper, we propose an efficient intra refreshment algorithm, which can achieve both the global optimization in determining the intra refreshing ratio and the local accuracy in selecting the proper macroblocks to be intra-coded before the current frame is actually encoded. Furthermore, incorporated with our proposed intra refreshment algorithm, an effective UEP scheme based on a dynamic FMO mapping mode is also proposed in this paper to better protect the intra macroblocks in a frame. The experimental results show that our joint intra refreshment and UEP algorithm can remarkably improve the reconstructed video quality in the packet lossy network.

### **MP1-P3.28 COST-EFFECTIVE FRAME-LAYER H.264 RATE CONTROL FOR LOW BIT RATE VIDEO**

*Changhyun Lee, Yonsei University, Republic of Korea; Seongjoo Lee, Sejong University, Republic of Korea; Yunje Oh, Samsung Electronics Co., Ltd., Republic of Korea; Jaeseok Kim, Yonsei University, Republic of Korea*

The rate control is important to allocate bits efficiently for getting better performance, such as high quality, low fluctuation of quality and low mismatch between a target bit rate and an encoded bit rate. In this paper, we present an improved frame-layer H.264/AVC rate control scheme using enhanced complexity measure, scene-change detection, and quantization parameter adjustment for low bit rate video. Especially, our complexity measure contains not only the residual information but also the motion information. We use the motion vector difference encoded previously as motion information for low computation-load encoder. Experimental results show that Y-PSNR and bit-rate mismatch are remarkably improved and the fluctuation of Y-PSNR is similar as compared with the conventional H.264/AVC rate control scheme (JVT-G012).

### **MP1-P3.29 A FAST VIDEO MOTION ESTIMATION ALGORITHM FOR THE H.264 STANDARD**

*Panos Nasiopoulos, Matthias von dem Kneesebeck, University of British Columbia, Canada*

Video applications are becoming an essential component for mobile devices. H.264, the latest video-coding standard, shows significant potential in terms of bandwidth savings at the cost of substantially increased complexity compared to former standards. The computing power currently available on mobile devices is not sufficient to allow high quality real-time encoding using H.264. Our algorithm uses on average only 0.41% of the computational complexity of the full search method used by H.264, leading to a significant reduction in computational requirements and enabling real-time applications for mobile devices with the efficiency of H.264.

### **MP1-P3.30 A BLOCK-MATCHING BASED INTRA FRAME PREDICTION FOR H.264/AVC**

*Jiheng Yang, Baocai Yin, Yanfeng Sun, Nan Zhang, Beijing University of Technology, China*

In the latest video coding standard H.264, intra frame prediction is employed to reduce the blocks' spatial correlation. In this paper, a new algorithm is proposed to improve the performance of intra prediction of H.264/AVC. It changes mode 2 of H.264's standard prediction methods into a BMA (block-matching algorithm)-DC hybrid mode. Experiment results show that with the proposed algorithm, significant improvement of coding performance can be gained compared to the performance of H.264 algorithm.

### **MP1-P3.31 FAST INTER MODE DECISION BASED ON HIERARCHICAL HOMOGENEOUS DETECTION AND COST ANALYSIS FOR H.264/AVC CODERS**

*Hung-Ming Wang, Ji-Kun Lin, Jar-Ferr Yang, National Cheng Kung University, Taiwan*

In H.264 advanced video coding (AVC), variable block size motion estimation plays an important role in compression of interframes. In this paper, we propose a fast inter prediction algorithm based on hierarchical homogeneous detection and cost analysis to select the best mode effectively. For each macroblock, we first detect that whether the macroblock is spatial homogeneous or not. For the non spatial homogeneous macroblock, we then perform the 16x16 motion estimation and examine if the 16x16 block is temporal homogeneous or not. Once the homogeneous macroblock is detected in the above process, the best mode will be chosen as 16x16 mode. For the non-homogeneous macroblock, we then execute 8x8 motion estimation and analyze the cost of 8x8 mode and 16x16 mode for deciding the best inter mode should be 16x16 mode or any other mode. The process for searching the best 8x8 block subtype is similar to the process for macroblocks. Finally, the best inter mode is decided by selecting the inter mode with least cost from the candidate modes. Experimental results show that our proposed algorithm can save about 32~54% computation time without introducing any noticeable performance degradation.

**MP1-P4 Coding and Compression (Poster)**

Time: Monday, July 10, 13:30 - 14:30

Place: Toronto III

Chair: Shahram Shirani, McMaster University

**MP1-P4.32 A FLEXIBLE CONTENT-BASED APPROACH TO ADAPTIVE IMAGE COMPRESSION**

*Alexander Wong, William Bishop, University of Waterloo, Canada*

Recent research in image compression has focused on lossy compression algorithms. However, the baseline implementations of such algorithms generally use a universal quantization process that results in poor image quality for certain types of images, particularly mixed-content images. This paper addresses this image quality issue by presenting a new algorithm that provides flexible and customizable image quality preservation by introducing an adaptive thresholding and quantization process based on content information such as edge and texture characteristics from the actual image. The algorithm is designed to improve visual quality based on the human vision system. Experimental results from the compression of various test images show noticeable improvements both quantitatively and qualitatively relative to baseline implementations as well as other adaptive techniques.

**MP1-P4.33 WAVELET-FUZZY-STOCHASTIC KALMAN FILTERING FOR IMAGE COMPRESSION**

*Nastoo Avessta, University of Turku, Finland*

This paper presents a novel fuzzy stochastic Kalman filter for compression of digital images. In particular, it is shown that the state evolution of the synthesis coefficients of any Discrete Wavelet Transform (DWT), in presence of coding degradation, may be described fuzzily. The novelty of this description is that, unlike other fuzzy based methods, it does not require a predefined membership measure. The fuzzy representation is further characterized by a stochastic nominal value and an interval of uncertainty. It is shown that such a framework allows for an efficient description of the underlying signals, particularly for images.

**MP1-P4.34 DIRECTIONAL DISCRETE COSINE TRANSFORMS FOR IMAGE CODING**

*Bing Zeng, Jingjing Fu, Hong Kong University of Science and Technology, Hong Kong SAR of China*

Nearly all block-based transform schemes for image and video coding developed so far choose the 2-D discrete cosine transform (DCT) of a square block shape. With almost no exception, this conventional DCT is implemented separately through two 1-D transforms, one along the vertical direction and another along the horizontal direction. In this paper, we develop a new block-based DCT framework in which the first transform may follow a direction other than the vertical or horizontal one, while the second transform is arranged to be a horizontal one. Compared to the conventional DCT, the resulting directional DCT framework is able to provide a better coding performance for image blocks that contain directional edges – a popular scenario in many image signals. By choosing the best from all directional DCT's (including the conventional DCT as a special case) for each image block, we will demonstrate that the rate-distortion coding performance can be improved remarkably.

**MP1-P4.35 IMPROVED METHODS FOR CORRELATING TRANSFORMS IN MULTIPLE DESCRIPTION CODING SYSTEMS**

*Claus Bauer, Corey Cheng, Dolby Laboratories, United States*

Multiple description has been widely investigated as a quality of service technology for the delivery of multimedia information over lossy channels. This paper presents generalized methods for transform based multiple description that do not introduce additional distortion to the transformed information. Previous noiseless methods place restrictions on how the information may be quantized prior to correlation. Typically, the information must be quantized with quasi-uniform quantizers which all have the same quantization step size. This paper expands these methods so that the information may be quantized with any quantizer, and different parts of the information may be quantized with different quantization step sizes. For specific correlating transforms, we propose a computationally efficient correlation transform that makes use of the Hadamard matrix.

**MP1-P4.36 A FAST MATCHING PURSUITS ALGORITHM USING SUB-BAND DECOMPOSITION OF VIDEO SIGNALS**

*Kousuke Imamura, Yoshimasa Koba, Hideo Hashimoto, Kanazawa University, Japan*

A high-efficiency video coding method using matching pursuits, which is a waveform coding technique, has recently been proposed. In this method, the motion compensated prediction error image is encoded by applying matching pursuits. In the present paper, we propose a matching pursuits coding method that encodes the sub-band images derived from motion compensated prediction error images. The complexity of the proposed method is reduced compared to the full-band matching pursuits because of lighter inner product computation due to reduction of both the resolution of the sub-band image and the basis function length of the dictionary. We evaluated the coding performance and computational complexity of the proposed method via computer simulations.

**MP1-P4.37 MULTI-RATE, DYNAMIC AND COMPLIANT REGION OF INTEREST CODING FOR JPEG2000**

*Jun Hou, Xiangzhong Fang, Jiliang Li, Haibin Yin, Songyu Yu, Shanghai Jiao Tong University, China*

A method is proposed to encode multiple regions of interest(ROI) in JPEG2000 image. It rearranges truncation point for every codeblock in each layer. It assigns higher bitrate to ROI and lower bitrate to non-ROI and combines them to codestream. The proposed strategy produces a fully compliant JPEG2000 codestream. It allows transmission of different ROIs with different priorities and supports dynamic delineation and prioritization of them. Experimental results demonstrating the validity of the proposed approach are presented.

**MP1-P4.38 NOVEL PROGRESSIVE REGION OF INTEREST IMAGE CODING BASED ON MATCHING PURSUITS**

*Abbas Ebrahimi-Moghadam, Shahram Shirani, McMaster University, Canada*

A progressive and scalable, region of interest (ROI) image coding scheme based on matching pursuits (MP) is presented. Matching pursuit is a multi-resolutional signal analysis tool and can be employed in order to progressively refine the quality of a set of selected regions of an image up to a specific grade. The computational complexity of this analysis method can be reduced by decreasing the size of MP dictionary. Thus, the proposed method provides a trade off between complexity, rate, and quality. By the suggested scheme, regions of an image with higher receiver's priority are refined in an interactive manner. The transmitter sends an initial coarse version of the image. Then, the receiver transmits its preferred ROI parameters. Afterwards, the reconstructed image is refined according to the ROI parameters, in a progressive way.

**MP1-P4.39 A 3D SPATIO-TEMPORAL MOTION ESTIMATION ALGORITHM FOR VIDEO CODING**

*Gwo Giun Lee, Ming-Jiun Wang, He-Yuan Lin, Drew Wei-Chi Su, Bo-Yun Lin, National Cheng Kung University, Taiwan*

This paper presents a new spatio-temporal motion estimation algorithm for video coding. The algorithm is based on optimization theory and consists of the strategies including 3D spatio-temporal motion vector prediction, modified one-at-a-time search scheme, and multiple update paths. The simulation results indicate our algorithm is better than other recently proposed ones under the same computational budget and is very close to full search. The low-cost feature and regular demand of computational resource make our algorithm suitable for VLSI implementation. The algorithm also makes single chip solution for high-definition coding feasible.

**MP1-P4.40 AN EFFECTIVE MODE DECISION SCHEME IN MACROBLOCK-BASED PFGS**

*Tiantian Sun, Chinese Academy of Sciences, China; Jizheng Xu, Feng Wu, Microsoft Research Asia, China; Wen Gao, Chinese Academy of Sciences, China*

The Progressive Fine Granularity Scalable (PFGS) coding is a promising technique for streaming video applications. The macroblock-based PFGS coding proposes several INTER modes for the enhancement macroblock coding and provides a simple method to determine a coding mode for each macroblock. Because it does not consider the drifting error possibly occurs in decoder, the mode decision method is not effective. In this paper, we present a mode decision scheme to make a trade-off between the low drifting error and high coding efficiency, by adding a middle bit-rate loop in encoder to estimate the possible drifting error in decoder and using a R-D model to estimate the improvement of coding efficiency by selecting high quality reference. Then a nearly optimal coding mode for each enhancement macroblock can be easily determined. The experimental results show that the proposed scheme can effectively reduce the drifting error at low bit rates, while maintaining improved coding efficiency at high bit rates.

## **MP1-P4.41 OPTIMIZATION OF MATCHING PURSUIT ENCODER BASED ON ANALYTICAL APPROXIMATION OF MATCHING PURSUIT DISTORTION**

*Alireza Shoa, Shahram Shirani, McMaster University, Canada*

Distortion of matching pursuit is calculated in terms of matching pursuit encoder parameters for uniformly distributed signals and dictionaries. Then, the MP encoder is optimized using the analytically derived approximation for MP distortion. Our simulation results show that this optimized MP encoder exhibits optimum performance for nonuniform signal and dictionary distributions as well.

### **MP2-L1: Monday, July 10, 14:30 - 15:50**

#### **MP2-L1 Object Tracking, Indexing and Searching (Lecture)**

**Time:** Monday, July 10, 14:30 - 15:50

**Place:** Carmichael

**Chair:** Stuart Schwartz, Princeton University

14:30 - 14:50

#### **MP2-L1.1 OBJECT TRACKING IN COMPRESSED VIDEO WITH CONFIDENCE MEASURES**

*Lan Dong, Princeton University, United States; Imad Zoghlami, Siemens Corporate Research, Inc., United States; Stuart C. Schwartz, Princeton University, United States*

In this paper, a novel robust tracking algorithm in compressed video is proposed. Within the framework of video compression standards, we consider how to accurately estimate motion of an object by utilizing motion vectors available in compressed video together with derived confidence measures. These confidence measures are based on DCT coefficients, spatial continuity of motion and texture measure of the object. We perform tracking directly on the compressed data and also consider tracking of an object with image scale change. In order to achieve robust tracking, we develop a system which enables us to detect object appearance change such as illumination change and occlusion by exploring the confidence measures derived above. Preliminary results indicate that our tracking algorithm works well with a variety of video sequences.

14:50 - 15:10

#### **MP2-L1.2 INTEGRATION OF BACKGROUND MODELING AND OBJECT TRACKING**

*Yu-Ting Chen, Chu-Song Chen, Academia Sinica, Taiwan; Yi-Ping Hung, National Taiwan University, Taiwan*

Background model and tracking became critical components for many vision-based applications. Typically, background modeling and object tracking are mutually independent in many approaches. In this paper, we adopt a probabilistic framework that uses particle filtering to integrate these two approaches, and the observation model is measured by Bhattacharyya distance. Experimental results and quantitative evaluations show that the proposed integration framework is effective for moving object detection.

15:10 - 15:30

#### **MP2-L1.3 A COLLECTION-ORIENTED METADATA FRAMEWORK FOR DIGITAL IMAGES**

*William Ku, Mohan S. Kankanhalli, National University of Singapore, Singapore; Joo-Hwee Lim, Institute for Infocomm Research, Singapore*

A digital photo can “tell a thousand words” through the use of its metadata and as it is usually part of a collection, metadata management, reuse, propagation & inference could be achieved via its association with a collection. However, there is not much work on metadata management, reuse, propagation & inference, particularly on a group basis. In this paper, we proposed a collection-oriented metadata framework which provides a basis for metadata management, reuse, propagation & inference and demonstrated the utility of such a framework.

15:30 - 15:50

**MP2-L1.4 EXPLORING AUTOMATIC QUERY REFINEMENT FOR TEXT-BASED VIDEO RETRIEVAL**

*Timo Volkmer, RMIT University, Australia; Apostol (Paul) Natsev, IBM T. J. Watson Research Center, United States*

Text-based search using video speech transcripts is a popular approach for granular video retrieval at the shot or story level. However, misalignment of speech and visual tracks, speech transcription errors, and other characteristics of video content pose unique challenges for this video retrieval approach. In this paper, we explore several automatic query refinement methods to address these issues. We consider two query expansion methods based on pseudo-relevance feedback and one query refinement method based on semantic text annotation. We evaluate these approaches in the context of the TRECVID 2005 Video Retrieval Benchmark using a baseline approach without any refinement. To improve robustness, we also consider a query-independent fusion approach. We show that this combined approach can outperform the baseline for most query topics, with improvements of up to 40%. We also show that query-dependent fusion approaches can potentially improve the results further, leading to 18-75% gains when tuned with optimal fusion parameters.

**MP2-L2: Monday, July 10, 14:30 - 15:50**

**MP2-L2 Perceptual Processing II (Special Session)**

Time: Monday, July 10, 14:30 - 15:50

Place: Johnston

Co-Chairs: Xiaokang Yang, Shanghai Jiao Tong University, Weisi Lin, Institute for Infocomm Research and Minoru Etoh, Multimedia Laboratories, NTT DoCoMo

14:30 - 14:50

**MP2-L2.1 PREFILTER CONTROL SCHEME FOR LOW BITRATE TV DISTRIBUTION**

*Ryoichi Kawada, Atsushi Koike, Yasuyuki Nakajima, KDDI R&D Laboratories, Inc, Japan*

In IP-based TV distribution, coding degradation is sometimes evident in critical scenes because the bit rate for compression is rather low. Prefiltering is an effective countermeasure since it replaces the coding noise with the degradation more difficult to detect visually, though it has the drawback that excessive smoothing might occur. This paper proposes a scene-adaptive method to control a prefilter separate from the encoder. By calculating block-wise motion-compensated predictive error variances and correlation coefficients, it estimates the coding noise as well as the potential improvement by prefiltering each frame, realizing a control scheme which performs prefiltering only when effective.

14:50 - 15:10

**MP2-L2.2 ON THE POTENTIAL OF INCORPORATING KNOWLEDGE OF HUMAN VISUAL ATTENTION INTO CBIR SYSTEMS**

*Oge Marques, Liam Mayron, Florida Atlantic University, United States; Gustavo Borba, Humberto Gamba, Universidade Tecnológica Federal do Paraná, Brazil*

Content-based image retrieval (CBIR) systems have been actively investigated over the past decade. Several existing CBIR prototypes claim to be designed based on perceptual characteristics of the human visual system, but even those who do are far from recognizing that they could benefit further by incorporating ongoing research in vision science. This paper explores the inclusion of human visual perception knowledge into the design and implementation of CBIR systems. Particularly, it addresses the latest developments in computational modeling of human visual attention. This fresh way of revisiting concepts in CBIR based on the latest findings and open questions in vision science research has the potential to overcome some of the challenges faced by CBIR systems.

15:10 - 15:30

### **MP2-L2.3 PERCEPTUAL SHARPNESS METRIC (PSM) FOR COMPRESSED VIDEO**

*Kai-Chieh Yang, Clark C. Guest, Pankaj Das, University of California, San Diego, United States*

Sharpness, one of the most effective factors in video quality assessment, usually dominates the first impression of the representation of the compressed video or image signals. In this paper, a new sharpness metric is presented. Without the original video sequence, this metric evaluates the level of sharpness of a compressed video sequence based on the presence of high frequency signals. Also, an attention module and several human visual factors are included in order to make the measuring results more correlated to human perception. Finally, psycovisual experiments show high correlation between the metric prediction and subjective ranking.

## **MP2-L3: Monday, July 10, 14:30 - 15:50**

### **MP2-L3 Content Understanding IV (Lecture)**

Time: Monday, July 10, 14:30 - 15:50

Place: Thomson

Chair: Jiebo Luo, Kodak Research Labs

14:30 - 14:50

### **MP2-L3.1 MIXED TYPE AUDIO CLASSIFICATION WITH SUPPORT VECTOR MACHINE**

*Lei Chen, Hong Kong University of Science and Technology, Hong Kong SAR of China; Sule Gündüz, Istanbul Technical University, Turkey; Tamer Özsu, University of Waterloo, Canada*

Content-based classification of audio data is an important problem for various applications such as overall analysis of audio-visual streams, boundary detection of video story segment, extraction of speech segments from video, and content-based video retrieval. Though the classification of audio into single type such as music, speech, environmental sound and silence is well studied, classification of mixed type audio data, such as clips having speech with music as background, is still considered a difficult problem. In this paper, we present a mixed type audio classification system based on Support Vector Machine (SVM). In order to capture characteristics of different types of audio data, besides selecting audio features, we also design four different representation formats for each feature. Our SVM-based audio classifier can classify audio data into five types: music, speech, environment sound, speech mixed with music, and music mixed with environment sound. The experimental results show that our system outperforms other classification systems using k Nearest Neighbor (k-NN), Neural Network (NN), and Naive Bayes (NB)

14:50 - 15:10

### **MP2-L3.2 USING SEMANTIC FEATURES FOR SCENE CLASSIFICATION: HOW GOOD DO THEY NEED TO BE?**

*Matthew Boutell, Rose-Hulman Institute of Technology, United States; Anustup Choudhury, University of Rochester, United States; Jiebo Luo, Eastman Kodak Company, United States; Christopher Brown, University of Rochester, United States*

Semantic scene classification is a useful, yet challenging problem in image understanding. Most existing systems are based on low-level features, such as color or texture, and succeed to some extent. Intuitively, semantic features, such as sky, water, or foliage, which can be detected automatically, should help close the so-called semantic gap and lead to higher scene classification accuracy. To answer the question of how accurate the detectors themselves need to be, we adopt a generally applicable scene classification scheme that combines semantic features and their spatial layout as encoded implicitly using a block-based method. Our scene classification results show that although our current detectors collectively are still inadequate to outperform low-level features under the same scheme, semantic features hold promise as simulated detectors can achieve superior classification accuracy once their own accuracies reach above a nontrivial 90%.



15:10 - 15:30

**MP2-L3.3 SIMILAR SEGMENT DETECTION FOR MUSIC STRUCTURE ANALYSIS VIA VITERBI ALGORITHM**

*Yu Shiu, University of Southern California, United States; Hong Jeong, Pohang University of Science and Technology, United States; C.-C. Jay Kuo, University of Southern California, United States*

The analysis of audio signals of songs of the verse-chorus form to reconstruct its original musical structures is investigated in this work. We first compute the similarity degree between any two measures in a song based on selected features and represent these numbers in a measure-based similarity matrix. Then, we study the similarity across a sequence of consecutive measures, which is revealed by straight segments in parallel with the diagonal line of the similarity matrix. Generally, chorus parts have higher similarity values while verse parts sometimes have lower similarity values. As a result, the verse parts are more difficult to detect. To tackle this problem systematically, the Viterbi Algorithm is adopted to find optimal paths in the lower-triangular similarity matrix, which represent repetitive segments of both choruses and verses. Finally, several post-processing steps are developed to decode the music structure into the verse, the chorus and other non-repetitive parts. Experimental results obtained from several musical audio data are shown to demonstrate the performance of the proposed method.

15:30 - 15:50

**MP2-L3.4 UTTERANCE-LEVEL EXTRACTIVE SUMMARIZATION OF OPEN-DOMAIN SPONTANEOUS CONVERSATIONS WITH RICH FEATURES**

*Xiaodan Zhu, Gerald Penn, University of Toronto, Canada*

To identify important utterances from open-domain spontaneous conversations, previous work has focused on using textual features that are extracted from transcripts, e.g., word frequencies and noun senses. In this paper, we summarize spontaneous conversations with features of a wide variety that have not been explored before. Experiments show that the use of speech-related features improves summarization performance. In addition, the effectiveness of individual features is examined and compared.

**MP2-L4: Monday, July 10, 14:30 - 15:50**

**MP2-L4 Coding Algorithms and Standards I (Lecture)**

Time: Monday, July 10, 14:30 - 15:50

Place: Toronto I

Chair: Mary Comer, Purdue University

14:30 - 14:50

**MP2-L4.1 NONLINEARLY-ADAPTED LAPPED TRANSFORMS FOR INTRA-FRAME CODING**

*Dan Lelescu, DoCoMo Communications Labs USA, United States*

The use of block transforms for coding intra-frames in video coding may preclude higher coding performance due to residual correlation across block boundaries and insufficient energy compaction, which translates into unrealized rate-distortion gains. Subjectively, the occurrence of blocking artifacts is common. Post-filters and lapped transforms offer good solutions to these problems. Lapped transforms offer a more general framework which can incorporate coordinated pre- and post-filtering operations. Most common are fixed lapped transforms (such as lapped orthogonal transforms), and also transforms with adaptive basis function length. In contrast, in this paper we determine a lapped transform that non-linearly adapts its basis functions to local image statistics and the quantization regime. This transform was incorporated into the H.264/AVC codec, and its performance evaluated. As a result, significant rate-distortion gains of up to 0.45 dB (average 0.35dB) PSNR were obtained compared to the H.264/AVC codec alone.

14:50 - 15:10

**MP2-L4.2 ANALYTICAL MODEL OF ZERO QUANTIZED DCT COEFFICIENTS FOR VIDEO ENCODER OPTIMIZATION**

*Hanli Wang, Sam Kwong, City University of Hong Kong, China; Chi-Wah Kok, Hong Kong Applied Science & Technology Research Institution Co. Ltd, China*

This paper proposes a novel analytical model to predict zero quantized DCT coefficients for fast video encoding. The dynamic range of quantized DCT coefficients are analyzed and a threshold scheme is derived to determine DCT and quantization computations to be skipped without video quality degradation. The proposed model is compared with other models in the literature. Experimental results demonstrate that the proposed analytical model can greatly reduce the computational complexity of video encoding without any performance degradation, and outperforms other models.

15:10 - 15:30

**MP2-L4.3 ADAPTIVELY SWITCHING BETWEEN DIRECTIONAL INTERPOLATION AND REGION MATCHING FOR SPATIAL ERROR CONCEALMENT BASED ON DCT COEFFICIENTS**

*Yan Chen, Oscar C. Au, Jiantao Zhou, Chi-Wang Ho, Hong Kong University of Science and Technology, China*

In this paper, a novel spatial error concealment algorithm, which adaptively switches between directional interpolation and region matching, is proposed. Different from the previous spatial error concealment methods, which just utilize smooth property, the algorithm exploits both smooth property and texture information to recover the lost blocks. Based on the DCT coefficients in the available neighboring MBs, the algorithm automatically analyzes whether the MB is “smooth-like” or “texture-like” and adaptively select directional interpolation or region matching to recover the lost MB. The proposed algorithm has been evaluated on H.264 reference software JM 9.0. The experimental results demonstrate that the proposed method can achieve better PSNR performance and visual quality, compared with weighted pixel average (WPA) which is adopted in H.264, directional interpolation-only and region matching-only.

15:30 - 15:50

**MP2-L4.4 BLOCK-BASED FAST COMPRESSION FOR COMPOUND IMAGES**

*Wenpeng Ding, Dong Liu, University of Science and Technology of China, China; Yuwen He, Feng Wu, Microsoft Research Asia, China*

**Abstract**—This paper presents a novel block-based fast compression (BFC) algorithm for compound images that contain graphics, text and natural images. The images are divided to blocks, which are classified into four different types – smooth blocks, text blocks, hybrid blocks and picture blocks with a fast and effective block-based classification algorithm. Four different coding algorithms are carefully designed for each block type according to their different statistical properties to maximize the compression performance. Simulations show that the BFC algorithm we propose has much lower complexity than DjVu with significant better visual quality at high bit rate, and it also outperforms the popular lossy image coding method JPEG.

**MP2-L5: Monday, July 10, 14:30 - 15:50**

**MP2-L5 P2P Multimedia II (Special Session)**

**Time:** Monday, July 10, 14:30 - 15:50

**Place:** Toronto II

**Co-Chairs:** Sherman Shen, University of Waterloo and Madjid Merabti, Liverpool John Moores University

14:30 - 14:50

**MP2-L5.1 A CROSS-LAYERED PEER-TO-PEER ARCHITECTURE FOR WIRELESS MOBILE NETWORKS**

*Mohammad Akon, Sagar Naik, Ajit Singh, Xuemin Shen, University of Waterloo, Canada*

In this paper, we propose a novel peer-to-peer architecture for wireless mobile networks where a cross-layered gossip-like protocol is the heart of the architecture. The goal of this architecture is to reduce the bandwidth consumption and at the same time, to provide more user participation flexibility. Simulation results are given to verify our goal.

14:50 - 15:10

**MP2-L5.2 ON THE DESIGN OF PREFETCHING STRATEGIES IN A PEER-DRIVEN VIDEO ON-DEMAND SYSTEM**

*Yanning Shen, Zhengye Liu, Shivendra Panwar, Keith Ross, Yao Wang, Polytechnic University, United States*

In this paper, we examine the prefetching strategies in a peerdriven video on-demand system. In our design, each video is encoded into multiple low bit-rate substreams and copies of the substreams are distributed to the participating peers. When a peer streams in a substream of rate  $r$ , it instead streams at rate  $r^{\wedge}$  where  $r^{\wedge} > r$ . In this manner, if one of the peer's suppliers disconnects, the client peer can tap the reservoir of prefetched bits while searching for a replacement server, thereby avoiding any glitches or reduced visual quality. We examine how to assign prefetching rates to each of substreams as a function of their importance. Our studies show that appropriate prefetching strategies can bring significant performance improvements for both multiple description and layered videos.

15:10 - 15:30

**MP2-L5.3 OFFERING DIFFERENTIATED SERVICES IN PEER-TO-PEER MULTIMEDIA MULTICAST**

*Kan-Leung Cheng, Xing Jin, S.-H. Gary Chan, Hong Kong University of Science and Technology, Hong Kong SAR of China*

In traditional overlay multicast network, all peers are treated as equal regardless of their importance and contribution to the network. In this paper, we consider that each user may have different cost function depending on his privilege, delay to source, available bandwidth, etc. We propose a fully distributed and scalable protocol to construct an overlay tree to minimize the overall cost of the users. It uses aggregation to account for the total cost of one's descendants, and reduces tree cost by a continuous improvement process. Through simulation, we show that our protocol converges reasonably fast. By comparing with other schemes, our resultant overlay tree offers differentiated services to users by appropriately taking into account individual user cost functions.

15:30 - 15:50

**MP2-L5.4 INFLUENCE OF GRAPH PROPERTIES OF PEER-TO-PEER TOPOLOGIES ON VIDEO STREAMING WITH NETWORK CHANNEL CODING**

*Shirish Karande, Hayder Radha, Michigan State University, United States*

Network Channel Coding (NCC) distributes channel coding functions over network nodes participating in common or diverse communication sessions. A particular case of NCC is Network Embedded FEC (NEF) [1]-[2], which has been shown to exhibit significant improvements in the performance of video streaming applications over multicast peer-to-peer (p2p) trees. The placement of NCC/NEF codecs and its utility in improving the throughput performance is in general a function of the underlying p2p graph topology. In this paper we consider two major forms of p2p topologies: (1) perfectly structured k-ary tree topologies that can be built from (virtually) ideal p2p graphs [3] and (2) unstructured random tree topologies where new nodes randomly join as children to any of the existing peers. The two topologies represent an optimal low diameter structured p2p topology and a trivial randomly evolving sub-optimal topology, respectively. In this paper, we show the impact of key graph parameters, such as the maximum node-degree  $k$  and minimum tree-height  $D$ , on the performance of NCC in terms of NEF throughput as well as video quality for both structured and unstructured topologies. The utility of NCC/NEF for low-degree and/or less structured p2p topologies is especially highlighted by demonstrating that, embedding of additional codecs can render the performance of less structured topologies or higher diameter topologies to be almost as good as that of the very well structured low diameter topologies. We also investigate the impact of the graph properties on the placement of the NEF codecs.

**MP2-L6 Media Signal Processing II (Lecture)**

Time: Monday, July 10, 14:30 - 15:50

Place: Varley

Chair: Steve Mann, University of Toronto

14:30 - 14:50

**MP2-L6.1 COST-EFFECTIVE SHARPENING OF SINGLE-SENSOR CAMERA IMAGES**

*Rastislav Lukac, Konstantinos N. Plataniotis, University of Toronto, Canada*

This paper presents a cost-effective image sharpening solution suitable for single-sensor digital cameras. The proposed solution enhances the structural content of the sensor image data captured using a Bayer color filter array (CFA). Subsequent demosaicking of the enhanced CFA image produces a visually-pleasing demosaicked image without the need for expensive sharpening in the RGB color domain. Simulation studies indicate that the proposed imaging pipeline, which employs an image sharpening step before the actual demosaicking module, yields excellent performance and outperforms the conventional pipeline (demosaicking followed by full-color image sharpening) in terms of both subjective and objective image quality measures.

14:50 - 15:10

**MP2-L6.2 PERCEPTUALLY-ADAPTIVE MOTION COMPENSATED TEMPORAL FILTERING**

*Dongdong Zhang, Xin Ma, Wenjun Zhang, Xiaokang Yang, Songyu Yu, Shanghai Jiao Tong University, China*

We propose a perceptually-adaptive motion compensated temporal filtering (MCTF) method to enhance the visual quality of 3D wavelet video coding schemes with spatial-domain MCTF. In our scheme, a spatio-temporal masking model in image domain is incorporated into the lifting structure of MCTF. The model is used to guide the motion search and the prediction step in MCTF so as to remove the visual redundancy in the video sequence. Experimental results show that the proposed scheme can significantly improve the visual quality of decoded video at different bitrates.

15:10 - 15:30

**MP2-L6.3 AN EDGE-BASED MEDIAN FILTERING ALGORITHM WITH CONSIDERATION OF MOTION VECTOR RELIABILITY FOR ADAPTIVE VIDEO DEINTERLACING**

*Qian Huang, Wen Gao, Debin Zhao, Qingming Huang, Chinese Academy of Sciences, China*

Because of its ability to preserve signal edges while filtering out impulsive noises, median filtering is widely used in signal processing applications, e.g. deinterlacing. An edge-based median filtering (EMF) algorithm is proposed for adaptive deinterlacing. A criterion of motion vector reliability (MVR) is also introduced for better interpolation. For each motion compensated block, two motion vectors between opposite-parity fields and one between same-parity fields are taken into account. Experiments show that the proposed MVR and EMF are both very efficient. Outputs of the proposed EMF are much more similar to original progressive videos than those of objectively best EMF methods nowadays, without obvious visual distortions. Finally, the proposed EMF and MVR criterion are shown to be suitable for texture-based adaptive deinterlacing.

15:30 - 15:50

**MP2-L6.4 AN ENCODER-EMBEDDED VIDEO DENOISING FILTER BASED ON THE TEMPORAL LMMSE ESTIMATOR**

*Liwei Guo, Oscar C. Au, Mengyao Ma, Zhiqin Liang, Hong Kong University of Science and Technology, Hong Kong SAR of China*

Noise not only degrades the visual quality of video contents, but also significantly affects the coding efficiency. Based on the temporal linear minimum mean square error (LMMSE) estimator, an innovative denoising filter is proposed in this paper. The proposed filter only requires simple operations manipulating on the individual residue coefficients and can be seamlessly integrated into video encoders. Compared to traditional filter-encoder cascaded scheme, embedding the proposed filter into the video encoder can save a large amount of computation. The experimental results show that with the proposed filter embedded, both the noise suppression capability and the coding efficiency of the video encoder can be dramatically improved. Furthermore, as a purely temporal filter, it can well preserve the fine details of video contents and satisfactory visual quality can be achieved.

**TA1-L1 Digital Watermarking, Data Hiding and Steganography I (Lecture)**

Time: Tuesday, July 11, 09:30 - 10:50

Place: Carmichael

Chair: Min Wu, University of Maryland, College Park

09:30 - 09:50

**TA1-L1.1 SELF-EMBEDDING DATA HIDING FOR NON-GAUSSIAN STATE-DEPENDENT CHANNELS: LAPLACIAN CASE**

*Oleksiy Koval, Sviatoslav Voloshynovskiy, Thierry Pun, University of Geneva, Switzerland*

In this paper, we consider the problem of optimal self-embedding Laplacian data hiding for the state-dependent channels. In particular, we propose to decompose the Laplacian data using the paradigm of parallel source splitting. Experimental validation confirms the efficiency of the proposed approach.

09:50 - 10:10

**TA1-L1.2 WATERMARK SYNCHRONIZATION FOR FEATURE-BASED EMBEDDING: APPLICATION TO SPEECH**

*David J. Coumou, Gaurav Sharma, University of Rochester, United States*

We propose a novel framework for synchronization in feature-based data embedding systems. The framework is tolerant to de-synchronizing errors in feature estimates, which have hitherto crippled feature-based embedding methods. The method uses a concatenated coding system comprising of an outer  $q$ -ary LDPC code and an inner insertion-deletion code to recover from both de-synchronization caused by feature estimation discrepancies between the transmitter and receiver; and errors in estimated symbols arising from other channel perturbations. We illustrate the framework in a speech watermarking application employing pitch modification for data-embedding. We show that the method indeed allows recovery of watermark data even in the presence of de-synchronization errors in the underlying pitch-based embedding. The resilience of the method is also demonstrated over channels employing low bit rate speech encoders.

10:10 - 10:30

**TA1-L1.3 LOSSLESS VISIBLE WATERMARKING**

*Shu-Kei Yip, Oscar C. Au, Chi-Wang Ho, Hoi-Ming Wong, Hong Kong University of Science and Technology, Hong Kong SAR of China*

The embedding distortion of visible watermarking is usually larger than that of invisible watermarking. In order to maintain the signal fidelity after the watermark extraction, “lossless” property is highlighted in the visible watermarking. In this paper, we propose two lossless visible watermarking algorithms, Pixel Value Matching Algorithm (PVMA) and Pixel Position Shift Algorithm (PPSA). PVMA uses the bijective intensity mapping function to watermark a visible logo whereas PPSA uses circular pixel shift to improve the visibility of the watermark in the high variance region. For the application of medical and military, as they are sensitive to distortion, PVMA and PPSA can be used to insert a visible logo to prevent unauthorized use.

10:30 - 10:50

**TA1-L1.4 BLOCK-BASED LOSSLESS DATA HIDING IN DELTA DOMAIN**

*Shu-Kei Yip, Oscar C. Au, Hoi-Ming Wong, Chi-Wang Ho, Hong Kong University of Science and Technology, Hong Kong SAR of China*

Digital watermarking is one of the ways to prove the ownership and the authenticity of the media. However, some applications, such as medical and military, are sensitive to distortion, this highlights the needs of lossless watermarking. In this paper, we propose a new lossless data hiding algorithm in delta domain. A MSE discount is obtained by using checkerboard-pattern watermark sequences. The PSNR between the watermarked image and the original image is high and there is no “salt-and-peppers” artifact. The proposed algorithm can be extended to withstand the JPEG attack.

**TA1-L2      Multi-User Collaboration and Cross Layer Optimization in Wireless Multimedia Communication I**  
(Special Session)

Time:            Tuesday, July 11, 09:30 - 10:50

Place:           Johnston

Co-Chairs:    Zhu Li, Motorola Labs, Pascal Frossard, EPFL and Aggelos Katsaggelos, Northwestern University

09:30 - 09:50

**TA1-L2.1      RECEIVER-BASED OPTIMIZATION FOR VIDEO DELIVERY OVER WIRELESS LINKS**

*Carri Chan, Stanford University, United States; John Apostolopoulos, Hewlett-Packard Laboratories, United States; Yan Li, Qualcomm, United States; Nick Bambos, Stanford University, United States*

We consider transfer of video frames over a time-varying wireless channel. When the channel is good, the transmitter can send frames at a higher rate than the receiver can consume them via playout. In that case, we introduce the idea of admitting new frames even when the receiver buffer is full, by selectively evicting frames already in the buffer; we can also control the playout rate, so as to optimize the tradeoff between video distortion and the time to freeze when the channel turns bad and frames arrive at a lower rate than should be played out. The decision/control problem of whether to admit a new frame, which already stored one to evict to accommodate the new one, and at what rate to play out frames is formulated within a dynamic programming framework, and an interesting connection to the Knapsack problem is made. Application of the idea in a relevant simple system shows significant performance gains, indicating that it is a promising approach for improving video delivery performance over challenging wireless channels.

09:50 - 10:10

**TA1-L2.2      COSMOS: PEER-TO-PEER COLLABORATIVE STREAMING AMONG MOBILES**

*Man-Fung Leung, Shueng-Han Gary Chan, Oscar C. Au, Hong Kong University of Science and Technology, Hong Kong SAR of China*

In traditional mobile streaming networks such as 3G cellular networks, all users pull streams from a server. Such pull model leads to high streaming cost and problem in system scalability. In this paper, we propose and investigate a scalable and cost-effective protocol to distribute multimedia content to mobiles in a peer-to-peer manner. Our protocol, termed Collaborative Streaming among Mobiles (COSMOS), makes use of multiple description coding (MDC) and data sharing to achieve high performance. In COSMOS, only a few peers pull video descriptions through a telecommunication channel. Using a free broadcast channel (such as Wi-Fi and bluetooth), they share the descriptions to nearby neighbors in an ad-hoc manner. This way reduces greatly the telecommunication cost and cellular bandwidth requirement. As video descriptions are supplied by multiple peers, COSMOS is robust to peer failure. Since broadcasting is used to distribute video data, the protocol is highly scalable to large number of users. By taking turns to pull descriptions, we show through simulation that peers can effectively share, and hence substantially reduce, streaming cost. As peers can often obtain a number of descriptions from nearby neighbors, they enjoy lower delay as compared to a recent scheme CHUM.

10:10 - 10:30

**TA1-L2.3      ANALYSIS OF MULTI-USER CONGESTION CONTROL FOR VIDEO STREAMING OVER WIRELESS NETWORKS**

*Xiaoqing Zhu, Bernd Girod, Stanford University, United States*

When multiple video sources are live-encoded and transmitted over a common wireless network, each stream needs to adapt its encoding parameters to wireless channel fluctuations, so as to avoid congesting the network. We present a stochastic system model for analyzing multi-user congestion control for live video coding and streaming over a wireless network. Variations in video content complexities and wireless channel conditions are modeled as independent Markov processes, which jointly determine the bottleneck queue size of each stream. Interaction among multiple users are captured by a simple model of random traffic contention. Using the model, we investigate two distributed congestion control policies: an approach based on stochastic dynamic programming (SDP) and a greedy heuristic. Compared to fixed-quality coding with no congestion control, performance gains in the range of 0.5-1.3 dB in average video quality are reported for the optimized schemes from simulation results.

10:30 - 10:50

**TA1-L2.4 GAME-THEORETIC PARADIGM FOR RESOURCE MANAGEMENT IN SPECTRUM AGILE WIRELESS NETWORKS**

*Fangwen Fu, Ahmad Reza Fattahi, Mihaela van der Schaar, University of California, Los Angeles, United States*

We propose a new way of architecting the wireless multimedia communications systems by jointly optimizing the protocol stack at each station and the resource exchanges among stations. We model wireless stations as rational players competing for available wireless resources in a dynamic repeated game. We investigate and quantify the system performance and the impact of different cross-layer strategies deployed by wireless stations onto their own performance as well as the competing station performance. We show through simulations that the proposed game-theoretic resource management outperforms alternative techniques such as air-fair time and equal time resource allocation in terms of the total system utility.

**TA1-L3: Tuesday, July 11, 09:30 - 10:50**

**TA1-L3 Event and Activity Recognition I (Lecture)**

Time: Tuesday, July 11, 09:30 - 10:50

Place: Thomson

Chair: Noboru Babaguchi, Osaka University

09:30 - 09:50

**TA1-L3.1 IDENTIFY SPORTS VIDEO SHOTS WITH “HAPPY” OR “SAD” EMOTIONS**

*Jinjun Wang, Institute for Infocomm Research, Singapore; Engsiong Chng, Nanyang Technological University, Singapore; Changsheng Xu, Institute for Infocomm Research, Singapore; Hanqing Lu, Xiaofeng Tong, Chinese Academy of Sciences, China*

Semantic video content extraction and selection are critical steps in sports video analysis and editing. The identification of video segments can be from various semantic perspectives, e.g. certain event, player or emotional state. In this paper, we examined the possibility of automatically identifying shots with “happy” or “sad” emotion from broadcast sports video. Our proposed model first performs the sports highlight extraction to obtain candidate shots that possibly contain emotion information and then classifies these shots into either “happy” or “sad” emotion groups using Hidden Markov Model based method. The final experimental results are satisfactory.

09:50 - 10:10

**TA1-L3.2 VISUAL EVENT DETECTION USING MULTI-DIMENSIONAL CONCEPT DYNAMICS**

*Shahram Ebadollahi, Lexing Xie, IBM T. J. Watson Research Center, United States; Shih-Fu Chang, Columbia University, United States; John R. Smith, IBM T. J. Watson Research Center, United States*

A novel framework is introduced for visual event detection. Visual events are viewed as stochastic temporal processes in the semantic concept space. In this concept-centered approach to visual event modeling, the dynamic pattern of an event is modeled through the collective evolution patterns of the individual semantic concepts in the course of the visual event. Video clips containing different events are classified by employing information about how well their dynamics in the direction of each semantic concept matches those of a given event. Results indicate that such a data-driven statistical approach is in fact effective in detecting different visual events such as “exiting car”, “riot”, and “airplane flying”.



10:10 - 10:30

**TA1-L3.3 WHERE AM I? SCENE RECOGNITION FOR MOBILE ROBOTS USING AUDIO FEATURES**

*Selina Chu, Shrikanth Narayanan, C.-C. Jay Kuo, Maja Mataric, University of Southern California, United States*

Automatic recognition of unstructured environments is an important problem for mobile robots. We focus on using audio features to recognize different auditory environments, where they are characterized by different types of sounds. The use of audio information provides a complementary means of scene recognition that can effectively augment visual information. In particular, audio can be used toward both the analysis and characterization of the environment at a higher level of abstraction. We begin our investigation of recognizing different auditory environments with the audio information. In this paper, we utilize low-level audio features from a mobile robot and investigate using high-level features based on spectral analysis for scene characterization, and a recognition system was built to discriminate between different environments based on these audio features found.

10:30 - 10:50

**TA1-L3.4 TV VIEWING INTERVAL ESTIMATION FOR PERSONAL PREFERENCE ACQUISITION**

*Hiroaki Tanimoto, Naoko Nitta, Noboru Babaguchi, Osaka University, Japan*

The importance of personalized information services has been increasing. Description of personal preferences needs to be prepared beforehand to realize such services. We propose a system for automatically acquiring personal preferences from TV viewer's behaviors. Considering "when" a viewer is watching TV is highly related to the viewer's preferences, we focus on estimating the time interval during which a pre-registered viewer is watching TV. In this paper, we firstly describe the outline of the personal preference acquisition system, and address a method for estimating the TV viewing intervals based on the appearance of frontal faces. Experiments resulted in a precision rate of 97.1% and a recall rate of 70.6% on average for TV viewing interval estimation.

**TA1-L4: Tuesday, July 11, 09:30 - 10:50**

**TA1-L4 Audio and Speech Processing I (Lecture)**

Time: Tuesday, July 11, 09:30 - 10:50

Place: Toronto I

Chair: Zicheng Liu, Microsoft Research

09:30 - 09:50

**TA1-L4.1 COMBINED GESTURE-SPEECH ANALYSIS AND SPEECH DRIVEN GESTURE SYNTHESIS**

*Mehmet Emre Sargin, Koç University, Turkey; Oya Aran, Bogazici University, Turkey; Alexey Karpov, SPIRAS, Russian Federation; Ferda Ofli, Koç University, Turkey; Yelena Yasinnik, MIT, United States; Stephen Wilson, University College Dublin, Ireland; Engin Erzin, Yucel Yemez, A. Murat Tekalp, Koç University, Turkey*

Multimodal speech and speaker modeling and recognition are widely accepted as vital aspects of state of the art human-machine interaction systems. While correlations between speech and lip motion as well as speech and facial expressions are widely studied, relatively little work has been done to investigate the correlations between speech and gesture. Detection and modeling of head, hand and arm gestures of a speaker have been studied extensively and these gestures were shown to carry linguistic information. A typical example is the head gesture while saying "yes/no". In this study, correlation between gestures and speech is investigated. In speech signal analysis, keyword spotting and prosodic accent event detection has been performed. In gesture analysis, hand positions and parameters of global head motion are used as features. The detection of gestures is based on discrete pre-designated symbol sets, which are manually labeled during the training phase. The gesture-speech correlation is modelled by examining the co-occurring speech and gesture patterns. This correlation can be used to fuse gesture and speech modalities for edutainment applications (i.e. video games, 3-D animations) where natural gestures of talking avatars is animated from speech. A speechdriven gesture animation example has been implemented for demonstration.

09:50 - 10:10

**TA1-L4.2 PREDOMINANT VOCAL PITCH DETECTION IN POLYPHONIC MUSIC**

*Xi Shao, Institute for Infocomm Research / National University of Singapore, Singapore; Changsheng Xu, Institute for Infocomm Research, Singapore; Mohan.S. Kankanhalli, National University of Singapore, Singapore*

We present a novel method for predominant vocal pitch detection in two-channel polyphonic music. The proposed method contains two stages. In the first stage, we apply the Frequency Domain Independent Component Analysis (FD-ICA) for the two-channel polyphonic music to separate the vocal content from the background music. Considering the vocal singing voice and background music are two heterogeneous signals, we employ a statistical learning based method to solve the permutation inconsistency problem in FD-ICA. In the second stage, a noise insensitive vocal pitch detection method is proposed, which is robust to noise and errors introduced by the separation process in the first stage. The proposed method has been tested on the two-channel polyphonic music signals, and experimental results show promising performance.

10:10 - 10:30

**TA1-L4.3 EFFICIENT PARTIAL SPECTRUM RECONSTRUCTION USING AN ASYMMETRIC PQMF ALGORITHM FOR MPEG-CODED STEREO AUDIO**

*Wendong Huang, Ye Wang, National University of Singapore, Singapore*

This paper presents a novel algorithm of a scalable and efficient Pseudo-Quadrature Mirror Filters (PQMF), which is employed for partial decoding a single-layer audio bitstream such as MP3, typically coded in joint/MS mode. The proposed algorithm is a new extension to our previous work on scalable audio decoding and is designed for asymmetric partial spectrum reconstruction (APSR), where perceptually irrelevant computations are removed. Furthermore, an efficient up-sampling operation is introduced for right channel output. The slight distortions introduced by our simple up-sampling method are inaudible according to a set of perceptual evaluations. Simulation results show that 64.6% energy savings can be achieved for a typical configuration in comparison to the standard PQMF algorithm employed by MPEG-1 audio.

10:30 - 10:50

**TA1-L4.4 ACOUSTIC ECHO CANCELATION FOR HIGH NOISE ENVIRONMENTS**

*Amit Chhetri, Arizona State University, United States; Jack Stokes, Dinei Florêncio, Microsoft Research, United States*

We propose a dual-structured acoustic echo cancellation (AEC) architecture that improves both convergence time and misadjustment of a conventional adaptive subband AEC algorithm in high noise environments. In this architecture, one part performs smooth adaptation while the other part performs fast adaptation; a convergence detector is implemented for switching between the two adaptation schemes. We propose the momentum normalized least mean square algorithm for smooth adaptation and we implement the NLMS algorithm for fast adaptation. The current architecture provides up to 3-4 dB echo reduction improvement over a conventional adaptive subband AEC algorithm while minimizing near-end distortion and artifacts in the post-processed AEC output.

**TA1-L5 Peer-to-Peer Multimedia Streaming (Lecture)**

Time: Tuesday, July 11, 09:30 - 10:50

Place: Toronto II

Chair: Gary Chan, Hong Kong University of Science &amp; Technology

09:30 - 09:50

**TA1-L5.1 EVALUATION OF PRACTICAL SCALABILITY OF OVERLAY NETWORKS IN PROVIDING VIDEO-ON-DEMAND SERVICE***Jian-Guang Luo, Yun Tang, Jiang Zhang, Shi-Qiang Yang, Tsinghua University, China*

Recently, overlay networks have been proposed to address the problem of scalability in providing video-on-demand (VoD) service. However, from the perspective of service providing, their efficiency has not been carefully studied and still remains far from clear, especially considering the impacts of user interactivities and in the case of multiple files with different and varying popularities on sharing. Towards this end, in this paper, by analyzing more than 20,000,000 real workload traces, we first identify two practical factors which we believe have determinant impacts on the scalability: user interactivities and popularity differences among files. Then we further evaluate cache-and-relay (CR), a representative scheme of overlay networks, with the real workload traces. Simulation results show that CR only save about half of the server bandwidth even when there is no buffer constraint at clients, not so scalable as our original expectation.

09:50 - 10:10

**TA1-L5.2 SIM: SCALABLE ISLAND MULTICAST FOR PEER-TO-PEER MEDIA STREAMING***Xing Jin, Kan-Leung Cheng, S.-H. Gary Chan, Hong Kong University of Science and Technology, Hong Kong SAR of China*

Despite the fact that global multicast is still not possible in today's Internet, many local networks are already multicast-capable (the so-called multicast "islands"). However, most application-layer multicast (ALM) protocols for streaming has not taken advantage of the underlying IP multicast capability. As IP multicast is more efficient, it would be beneficial if ALM can take advantage of such capability in building overlay trees. In this paper, we propose a fully distributed protocol called Scalable Island Multicast (SIM), which effectively integrates IP multicast and ALM. Hosts in SIM first form an overlay tree using a scalable protocol. They then detect IP multicast islands and employ IP multicast whenever possible. Through simulations on Internet-like topologies, we show that SIM achieves much lower end-to-end delay and link stress as compared with traditional ALM protocols.

10:10 - 10:30

**TA1-L5.3 CONSTRUCTING ROBUST AND RESILIENT FRAMEWORK FOR COOPERATIVE VIDEO STREAMING***Shi Lu, Michael Lyu, The Chinese University of Hong Kong, Hong Kong SAR of China*

Peer-to-peer based streaming has been a promising solution for large-scale video broadcasting over the Internet. In a peer-to-peer video streaming framework, peers cooperate with each other for content distribution, so that the burden of the central server is greatly alleviated. Moreover, the peer-to-peer overlay is highly scalable to support a very large number of users. However, to support video streaming services, some strict performance issues need to be addressed, e.g., reliability, resilience and robustness to network dynamics. In this paper, we investigate several goals that a peer overlay should achieve in order to support good quality video streaming. We then describe our work on organizing peers into such a robust and resilient framework. We have implemented a fully functional video broadcasting system based on the proposed peer-to-peer infrastructure. The prototype system has been successfully deployed and tested upon Planet-lab with encouraging experimental results.

10:30 - 10:50

**TA1-L5.4 A PEER-TO-PEER OVERLAY NETWORK FOR REAL TIME VIDEO COMMUNICATION USING MULTIPLE PATHS**

*Andrea De Mauro, University of Illinois at Chicago, United States / Politecnico di Torino, Italy; Dan Schonfeld, University of Illinois at Chicago, United States; Claudio Casetti, Politecnico di Torino, Italy*

This paper presents a peer-to-peer (P2P) service for the transmission of real-time video content, exploiting the contemporary usage of multiple network paths over the current Internet. Before starting the transmission, the sender probes the available paths for their Round Trip Time and other parameters using simple ICMP packets. Then it chooses the set of paths to be used in order to maximize the expected Signal to Noise Ratio (SNR) while meeting the delay requirements of the streaming content. The video is encoded using a Multiple Description Coding (MDC) technique. Simulations show that the proposed algorithm always chooses one of the group of paths that yields the best SNR.

**TA1-L6: Tuesday, July 11, 09:30 - 10:50**

**TA1-L6 Interactive Multimedia Content Analysis (Special Session)**

Time: Tuesday, July 11, 09:30 - 10:50

Place: Varley

Co-Chairs: Xiang-Sheng Hua, Microsoft Research Asia and Qi Tian, UT San Antonio

09:30 - 09:50

**TA1-L6.1 VISUAL FEATURE SPACE ANALYSIS FOR UNSUPERVISED EFFECTIVENESS ESTIMATION AND FEATURE ENGINEERING**

*Tobias Schreck, Daniel Keim, University of Konstanz, Germany; Christian Panse, ETH Zurich, Switzerland*

The Feature Vector approach is one of the most popular schemes for managing multimedia data. For many data types such as audio, images, or 3D models, an abundance of different Feature Vector extractors are available. The automatic (unsupervised) identification of the best suited feature extractor for a given multimedia database is a difficult and largely unsolved problem. We here address the problem of comparative unsupervised feature space analysis. We propose two interactive approaches for the visual analysis of certain feature space characteristics contributing to estimated discrimination power provided in the respective feature spaces. We apply the approaches on a database of 3D objects represented in different feature spaces, and we experimentally show the methods to be useful (a) for unsupervised comparative estimation of discrimination power and (b) for visually analyzing important properties of the components (dimensions) of the respective feature spaces. The results of the analysis are useful for feature selection and engineering.

09:50 - 10:10

**TA1-L6.2 AN INERACTIVE REGION-BASED IMAGE CLUSTERING AND RETRIEVAL PLATFORM**

*Ying Liu, Xin Chen, Chengcui Zhang, Alan Sprague, University of Alabama at Birmingham, United States*

Content-based Image retrieval has become an important part of information retrieval technology. Images can be viewed as high dimensional data and are usually represented by their low-level features. How to effectively find the semantic meanings of images is a central challenge in the area. In this paper, we propose an interactive platform for region-based image clustering and retrieval. A Genetic Algorithm is used to perform the initial clustering. In order to further refine the clustering results, we adopt the maximum flow/minimum cut theorem from graph theory to do outlier/outlier group detection. Outlier detection can help identify misclustered image segments and is used to improve the quality of clusters in this paper. In the interactive retrieval phase, user feedback is used to dynamically locate candidate images from clusters and outliers/outlier groups. Through Relevance Feedback, more information is gathered and fed to the learning algorithm – One-class SVM. Experiments show the effectiveness of the proposed platform.

10:10 - 10:30

**TA1-L6.3 VIDEO ANNOTATION BY ACTIVE LEARNING AND SEMI-SUPERVISED ENSEMBLING**

*Yan Song, Guo-Jun Qi, University of Science and Technology of China, China; Xian-Sheng Hua, Microsoft Research Asia, China; Li-Rong Dai, Ren-Hua Wang, University of Science and Technology of China, China*

Supervised and semi-supervised learning are frequently applied methods to annotate videos by mapping low-level features into semantic concepts. Due to the large semantic gap, the main constraint of these methods is that the information contained in a limited-size labeled dataset can hardly represent the distributions of the semantic concepts. In this paper, we propose a novel semi-automatic video annotation framework, active learning with semi-supervised ensemble, which tries to tackle the disadvantages of current video annotation solutions. Firstly the initial training set is constructed based on distribution analysis of the entire video dataset. And then an active learning scheme is combined into a semi-supervised ensemble framework, which selects the samples to maximize the margin of the ensemble classifier based on both labeled and unlabeled data. Experimental results show that the proposed method performs superior to general semi-supervised learning algorithms and the typical active learning algorithms in terms of annotation accuracy and stability.

10:30 - 10:50

**TA1-L6.4 INTERACTIVE CONTENT-AWARE MUSIC BROWSING USING THE RADIO DRUM**

*Jennifer Murdoch, George Tzanetakis, University of Victoria, Canada*

Portable digital music players are becoming pervasive and the size of personal digital music collections has been steadily increasing (5-10 thousand tracks are common today). The emerging area of Music Information Retrieval (MIR) deals with all aspects of managing, analyzing and organizing music in digital formats. The majority of work in MIR follows a search/retrieval paradigm. More recently, the importance of browsing as an interaction paradigm has been realized and several novel interfaces have been proposed. In this paper, we describe a tangible interface for content-aware browsing of music collections. The Radio Drum is a gestural interface based on capacitance sensors that can detect the x,y,z positions of two drum sticks in a 3D volume. We describe two possible mappings that can be used for browsing music collections without relying on metadata. The first is an explicit mapping of tempo and beat strength, and the second is a music similarity space using audio feature extraction and a Self Organizing Map (SOM).

**TA1-P1: Tuesday, July 11, 09:30 - 10:30**

**TA1-P1 Content Understanding V (Poster)**

Time: Tuesday, July 11, 09:30 - 10:30

Place: Toronto III

Chair: Constantine Kotropoulos, Aristotle University of Thessaloniki

**TA1-P1.1 GENERATING EXPRESSIVE SUMMARIES FOR SPEECH AND MUSICAL AUDIO USING SELF-SIMILARITY CLUES**

*Mustafa Sert, Baskent University, Turkey; Buyurman Baykal, Adnan Yazici, Middle East Technical University, Turkey*

We present a novel algorithm for structural analysis of audio to detect repetitive patterns that are suitable for content-based audio information retrieval systems, since repetitive patterns can provide valuable information about the content of audio, such as a chorus or a concept. The Audio Spectrum Flatness (ASF) feature of the MPEG-7 standard, although not having been considered as much as other feature types, has been utilized and evaluated as the underlying feature set. Expressive summaries are chosen as the longest patterns by the k-means clustering algorithm. Proposed approach is evaluated on a test bed consisting of popular song and speech clips based on the ASF feature. The well known Mel Frequency Cepstral Coefficients (MFCCs) are also considered in the experiments for the evaluation of features. Experiments show that, all the repetitive patterns and their locations are obtained with the accuracy of 93% and 78% for music and speech, respectively.

**TA1-P1.2    A HIERARCHICAL APPROACH FOR MUSIC CHORD MODELING BASED ON THE ANALYSIS OF TONAL CHARACTERISTICS**

*Namunu C Maddage, Institute for Infocomm Research, Singapore; Mohan S. Kankanhalli, National University of Singapore, Singapore; Haizhou Li, Institute for Infocomm Research, Singapore*

This paper first discusses how the signal segmentation and tonal characteristics of music notes effect in music chord detection. Two approaches, pitch class profile approach and psycho-acoustical approach, which differently represent these tonal characteristics, are examined for chord detection. The analysis of the tonal characteristics reveals that not only the fundamental frequency of music note but also its harmonics and sub-harmonies in different octaves contribute for detecting related music chord. A hierarchical approach, which transforms the music chord tonal characteristics in each octave onto probabilistic space, is then proposed for modeling the music chord. Our experimental results show that detection of chord type, Major, Minor, Diminish, and Augmented, and individual chords, 12 chords per chord type, are improved with the proposed hierarchical chord modeling approach. Experimental results also reveal that the tempo proportional signal segmentation is more effective extracting tonal characteristics than using fixed length segmentation.

**TA1-P1.3    CONVERSATION SCENE ANALYSIS WITH DYNAMIC BAYESIAN NETWORK BASED ON VISUAL HEAD TRACKING**

*Kazuhiro Otsuka, Junji Yamato, NTT Communication Science Laboratories, Japan; Yoshinao Takemae, NTT Cyber Solutions Laboratories, Japan; Hiroshi Murase, Nagoya University, Japan*

A novel method based on a probabilistic model for conversation scene analysis is proposed that can infer conversation structure from video sequences of face-to-face communication. Conversation structure represents the type of conversation such as monologue or dialogue, and can indicate who is talking / listening to whom. This study assumes that the gaze directions of participants provide cues for discerning the conversation structure, and can be identified from head directions. For measuring head directions, the proposed method newly employs a visual head tracker based on Sparse-Template Condensation. The conversation model is built on a dynamic Bayesian network and is used to estimate the conversation structure and gaze directions from observed head directions and utterances. Visual tracking is conventionally thought to be less reliable than contact sensors, but experiments confirm that the proposed method achieves almost comparable performance in estimating gaze directions and conversation structure to a conventional sensor-based method.

**TA1-P1.4    SEGMENTATION AND RECOGNITION OF MEETING EVENTS USING A TWO-LAYERED HMM AND A COMBINED MLP-HMM APPROACH**

*Stephan Reiter, Björn Schuller, Gerhard Rigoll, Technische Universität München, Germany*

Automatic segmentation and classification of recorded meetings provides a basis that enables effective browsing and querying in a meeting archive. Yet, robustness of today's approaches is often not reliable enough. We therefore strive to improve on this task by introduction of a hybrid approach combining the discriminative abilities of artificial neural nets and warping capabilities of hidden markov models. Dividing the task into two layers and defining a proper set of individual actions helps to cope with the problem of lack of data and overcomes conventional single-layered approaches. Extensive test runs on the public M4 Scripted Meeting Corpus prove the great performance gain applying our suggested novel approach compared to other similar methods.

**TA1-P1.5    MUSICAL ONSET DETECTION BASED ON ADAPTIVE LINEAR PREDICTION**

*Wan-Chi Lee, C.-C. Jay Kuo, University of Southern California, United States*

A new musical onset detection technique based on adaptive linear prediction theory is proposed in this work. We decompose a music signal into multiple sub-bands, and then apply a forward linear prediction error filter (LPEF) to model the narrow-band signal in each band, respectively. To enhance the modeling performance, the coefficients of the LPEF are updated with the least-mean-squares (LMS) algorithm. Under this framework, the onset detection problem can be formulated as the peak-error location problem. Peak selection algorithms are applied to prediction errors to locate the onset time. It is shown by experimental results that the proposed algorithm outperforms several well known existing methods for onset detection.

**TA1-P1.6    VOICE ACTIVITY DETECTION WITH GENERALIZED GAMMA DISTRIBUTION**

*George Alamanidis, Constantine Kotropoulos, Aristotle University of Thessaloniki, Greece*

In this work, we model speech samples with the generalized Gamma distribution and evaluate the efficiency of such modelling for voice activity detection. Using a computationally inexpensive maximum likelihood approach, we employ the Bayesian Information Criterion for identifying the phoneme boundaries in noisy speech.

#### **TA1-P1.7 AUTOMATIC QUERY EXPANSION IN NEWS VIDEO RETRIEVAL**

*Yun Zhai, Jingen Liu, Mubarak Shah, University of Central Florida, United States*

In this paper, we present an integrated system for news video retrieval. The proposed system incorporates both speech and visual information in the search mechanisms. The initial search is based on the automatic speech recognition (ASR) transcript of video. Based on the relevant shots selected from the initial search round, keyword histograms are automatically generated for the refinement of the search query, such that the reformulated query fits better to the target topic. We have also developed an image-based refinement module, which uses the region analysis of the video key-frames. SR-tree like indexing structure is constructed for the region features, and the image-to-image similarity is computed using the Earth Mover's Distance. By performing a series of relevance feedback processes, the set of the true relevant shots is expanded significantly. The proposed system has been applied to a large open-benchmark news video dataset, and very satisfactory improvements have been obtained by applying the proposed automatic query expansion and the region-based refinement.

#### **TA1-P1.8 HIERARCHICAL SUMMARIZATION OF VIDEOS BY TREE-STRUCTURED VECTOR QUANTIZATION**

*Sergio Benini, Aldo Bianchetti, Riccardo Leonardi, Pierangelo Migliorati, University of Brescia, Italy*

Accurate grouping of video shots could lead to semantic indexing of video segments for content analysis and retrieval. This paper introduces a novel cluster analysis which, depending both on the video genre and the specific user needs, produces a hierarchical representation of the video only on a reduced number of significant summaries. An outlook on a possible implementation strategy is then suggested. Specifically, vector-quantization codebooks are used to represent the visual content and to cluster the shots with a similar chromatic consistency. The evaluation of the codebook distortion introduced in each cluster is used to stop the procedure on few levels, exploiting the dependency relationships between clusters. Finally, the user can navigate through summaries at each hierarchical level and then decide which level to adopt for eventual post-processing. The effectiveness of the proposed method is validated through a series of experiments on real visual-data excerpted from different kinds of programmes.

#### **TA1-P1.9 AUTOMATIC CLASSIFICATION OF PHOTOGRAPHS AND GRAPHICS**

*Yuanhao Chen, University of Science and Technology of China, China; Zhiwei Li, Mingjing Li, Wei-Ying Ma, Microsoft Research Asia, China*

In general, digital images can be classified into photographs and computer graphics. This taxonomy is very useful in many applications, such as web image search. However, there are no effective methods to perform this classification automatically. In this paper, we manage to solve this problem from two aspects. At first, we propose some novel low-level features that can reveal perceptual differences between photographs and graphics. Then, we adopt an effective algorithm to perform the classification. The experiments conducted on a large-scale image database indicate the effectiveness of our algorithm.

#### **TA1-P1.10 EFFICIENT SPEAKER DETECTION VIA TARGET DEPENDENT DATA REDUCTION**

*Upendra Chaudhari, Olivier Verscheure, Juan Huerta, Xiang Li, Ganesh Ramaswamy, Lisa Amini, IBM T. J. Watson Research Center, United States*

Systems designed to extract time-critical information from large volumes of unstructured data must include the ability, both from an architectural and algorithmic point of view, to filter out unimportant data that might otherwise overwhelm the available resources. This paper presents an approach for data filtering to reduce computation in the context of a distributed speech processing architecture designed to detect or identify speakers. Here, filtering means either dropping and ignoring data or passing it on for further processing. The goal of the paper is to show that when the filter is designed to select and pass on a subset of the input data that best preserves the ability to recognize a specific desired speaker, or group of speakers, a large percentage of the data can be ignored while being able to preserve most of the accuracy.



**TA1-P2 Object Recognition and Tracking (Poster)**

Time: Tuesday, July 11, 09:30 - 10:30

Place: Toronto III

Chair: Ioannis Pitas, Aristotle University of Thessaloniki

**TA1-P2.11 DISPARITY-BASED 3D FACE MODELING USING 3D DEFORMABLE FACIAL MASK FOR 3D FACE RECOGNITION**

*A-Nasser Ansari, Mohamed Abdel-Mottaleb, Mohammad H. Mahoor, University of Miami, United States*

We present an automatic disparity-based approach for 3D face modeling, from two frontal and one profile view stereo images, for 3D face recognition applications. Once the images are captured, the algorithm starts by extracting selected 2D facial features from one of the frontal views and computes a dense disparity map from the two frontal images. Using the extracted 2D features plus their corresponding disparities in the disparity map, we compute their 3D coordinates. We next align a low resolution 3D mesh model to the 3D features, re-project it's vertices on the frontal 2D image and adjust its profile line vertices using the profile view. We increase the resolutions of the resulting 2D model only at its center region to obtain a facial mask model covering distinctive features of the face. The computation of the 2D vertices coordinates with their disparities results in a deformed 3D model mask specific to a give subject's face. Application of the model in 3D face recognition validates the algorithm and shows a high recognition rate.

**TA1-P2.12 FUZZY SPATIAL RANKS FOR OBJECT RECOGNITION ACROSS ILLUMINATION CHANGES**

*Damien Muselet, Ludovic Macaire, UMR CNRS 8146, France*

In this paper, we propose an original scheme to retrieve among all the target images of a database, those which contain the same object as that represented by the query image, these images being acquired under different illumination conditions. Rather than considering the color vectors of the pixels to characterize the images, we propose to introduce and exploit the concept of spatial ranks of CCD sensor responses. Indeed, these values are preserved in case of illumination changes and they take into account both the colors of the pixels and the spatial interactions between them in the image. Since we can not determine these ranks from a color image, we propose to estimate their probabilities of occurrences thanks to fuzzy functions. These probabilities are used by our object recognition scheme whose effectiveness is assessed with a public database that contains images of objects acquired under different illuminations.

**TA1-P2.13 FAST VIDEO OBJECT SELECTION FOR INTERACTIVE TELEVISION**

*Rémi Trichet, Bernard Merialdo, Institut Eurecom, France*

In this paper, we study the problem of the fast selection of video objects, as an aid for the efficient semi-automatic annotation of video programs. In a regular system, the user has to draw a bounding box around the object, requiring at least two clicks on the image. We propose and experiment algorithms that allow the selection by indicating only one point inside the object, therefore requiring only one click. The problem is then to identify the correct bounding box. We use an attention model and a growing algorithm to construct the most plausible bounding box, based on the comparison of the interior, the border and the outside of the box. We present some experimentation that suggests that in many cases, our algorithm is able to propose a reasonable bounding box.

**TA1-P2.14 OBJECT RECOGNITION AND RECOVERY BY SKELETON GRAPH MATCHING**

*Lei He, Armstrong Atlantic State University, United States; Chia Han, William Wee, University of Cincinnati, United States*

This paper presents a robust and efficient skeleton-based graph matching method for object recognition and recovery applications. The novel feature is to unify both object recognition and recovery components into an image understanding system architecture, in which a complementary feedback structure can be incorporated to alleviate processing difficulties of each component alone. The idea is firstly to recognize the preliminary extracted object from a set of models using the new skeleton graph matching method, then to apply the a priori shape information of the identified model for accurate object recovery. The output of the system is the recognized and segmented object. The skeleton graph matching method is illustrated by recognizing a set of tool and animal silhouette examples with the presence of geometric transformations (translation, rotation, scaling, reflection), shape deformations and noise. Experiments of object recovery using MR knee images, have shown satisfactory results.

#### **TA1-P2.15 IMPROVED GRAPHICAL MODEL FOR AUDIOVISUAL OBJECT TRACKING**

*Hao Tang, Thomas S. Huang, University of Illinois at Urbana-Champaign, United States*

Object tracking plays an important role in multimedia surveillance systems, in which the major types of data are video and audio captured by cameras and microphone arrays. In this paper, we describe a systematic approach to audiovisual object tracking, originally proposed by Beal et al, based on graphical models that jointly combine audio and video variables under a single probabilistic framework. We seek to improve this approach through three aspects: First, we introduce background subtraction preprocessing of video data. Second, we modify the video model to exclude the background from being transformed. Third, we extend the joint model to a dynamic Bayes net. These improvements yield satisfactory results on single person tracking in a noisy outdoor environment with far-field background road traffic, and handle situations where the target is lost due to occlusions.

#### **TA1-P2.16 EFFICIENT OBJECT TRACKING USING CONTROL-BASED OBSERVER DESIGN**

*Wei Qu, Dan Schonfeld, University of Illinois at Chicago, United States*

Kernel-based tracking approaches have proven to be more efficient in computation compared to other tracking approaches such as particle filtering. However, existing kernel-based tracking approaches suffer from the well-known "singularity" problem. In this paper, we propose a novel object tracking framework to handle this problem by using a control-based observer design. Specifically, we formulate object tracking as a recursive inverse problem, thus unifying several approaches to video tracking, including kernel-based tracking, into a consistent theoretical framework. Moreover, we interpret the inverse equation as a measurement process and supplement it by introducing state dynamics as a constraint. The augmented recursive inverse equation forms a state-space model, which is solved by using a control-based optimal observer. By exploiting observability theory from control engineering, we extend the current approach to kernel-based tracking and provide explicit criteria for kernel design and dynamics evaluation. The tracking performance of our approach has been demonstrated on both synthetic and real-world video data.

#### **TA1-P2.17 PEOPLE IDENTIFICATION WITH LIMITED LABELS IN PRIVACY-PROTECTED VIDEO**

*Yi Chang, Rong Yan, Datong Chen, Jie Yang, Carnegie Mellon University, United States*

People identification is an essential task for video content analysis in surveillance systems. A good classifier, requires a large amount of training data, which may not be obtained in some scenario. In this paper, we propose an approach to augment insufficient training data with pairwise constraints. We show user study results that human subjects can perform reasonably well in labeling pairwise constraints from face masked images. We also present a new discriminative learning algorithm that can handle uncertainties in pairwise constraints. The new method provides a way to obtain high accuracy of people identification from limited labeled data with noisy pairwise constraints, and meanwhile minimize the risk of exposing people's identities.

#### **TA1-P2.18 COARSE-TO-FINE PEDESTRIAN LOCALIZATION AND SILHOUETTE EXTRACTION FOR THE GAIT CHALLENGE DATA SETS**

*Haiping Lu, Konstantinos N. Plataniotis, Anastasios (Tas) N. Venetsanopoulos, University of Toronto, Canada*

This paper presents a localized coarse-to-fine algorithm for efficient and accurate pedestrian localization and silhouette extraction for the Gait Challenge data sets. The coarse detection phase is simple and fast. It locates the target quickly based on temporal differences and some knowledge on the human target. Based on this coarse detection, the fine detection phase applies a robust background subtraction algorithm to the coarse target regions and the detection obtained is further processed to produce the final results. This algorithm has been tested on 285 outdoor sequences from the Gait Challenge data sets, with wide variety of capture conditions. The pedestrian targets are localized very well and silhouettes extracted resemble the manually labeled silhouettes closely.

#### **TA1-P2.19 A MUTUAL INFORMATION BASED FACE CLUSTERING ALGORITHM FOR MOVIES**

*Nicholas Vretos, Vassilios Solachidis, Ioannis Pitas, Aristotle University of Thessaloniki, Greece*

In this paper a new approach for face clustering is developed. Mutual information and joint entropy are exploited in order to create a metric for the clustering process. The way the joint entropy and the mutual information are calculated gives some interesting properties to the aforementioned metric, which guarantees some robustness against standard noisy transformation such as scaling, cropping and pose changes. A slight preprocessing of the input face images is done in order to undertake problems that arise from detector's output known errors.

## **TA1-P2.20 OPTIMIZED CHAMFER MATCHING FOR SNAKE-BASED IMAGE CONTOUR REPRESENTATIONS**

*András Hajdu, Athanasios Roubies, Ioannis Pitas, Aristotle University of Thessaloniki, Greece*

In this paper we present a novel method on how to take advantage of the snake representation of target objects, when doing chamfer matching for detection/recognition purposes. In this case several time-consuming steps of classic chamfer matching approaches can be simplified. Moreover, we investigate the possibility of involving less pixels from both the target and template object to speed up computations. We introduce an optimization method for such an object reduction, which is valid also in the general application scheme of chamfer matching. Finally, we present our experimental results regarding human body detection.

### **TA1-P3: Tuesday, July 11, 09:30 - 10:30**

#### **TA1-P3 Other Applications (Poster)**

Time: Tuesday, July 11, 09:30 - 10:30

Place: Toronto III

Chair: Mubarak Shah, University of Central Florida

#### **TA1-P3.21 REAL-TIME VIDEO INTELLIGENT SURVEILLANCE SYSTEM**

*Weidong Zhang, Feng Chen, Wenli Xu, Enwei Zhang, Tsinghua University, China*

With the rapid development of hardware equipments, it is now economically and technically feasible to build a video surveillance system. This paper presents the system architecture of VISS, a Video Intelligent Surveillance System deployed in parking lots. In VISS we adopt robust moving object detecting and tracking algorithm, and we present a novel activity recognition framework based on Layer Hidden Semi-Markov Model (LHSM) which is used for modeling activities. The experimental results on real-time video show that the system is effective and robust in complex activity recognition.

#### **TA1-P3.22 GPCD:GRID-BASED PREDICTIVE COLLISION DETECTION FOR LARGE-SCALE ENVIRONMENTS IN COMPUTER GAMES**

*Zhiwen Yu, Hau-San Wong, City University of Hong Kong, Hong Kong SAR of China*

Given a time horizon parameter  $h$  and an object set  $O$ , predictive collision detection finds all the object pairs  $\langle o_i, o_j \rangle$ ,  $t_i >$  which will collide in the future time interval  $[t, t + h]$  (where  $1 < i, j < n$ ,  $t_i$  belongs to  $[t, t + h]$ ). Although there are a number of state-of-the-art approaches to solve collision detection problems, predictive collision detection is addressed for the first time. In this paper, we propose a grid-based predictive collision detection algorithm (GPCD), which is a general technique for the efficient detection of the collision of object pairs in a future time interval. GPCD first determines a candidate list which stores the object pairs having a non-zero probability to collide in a future time. Then, GPCD achieves low running time based on two pruning strategies: (i) space intersection test and (ii) time intersection test. These two pruning strategies eliminate most of the false collision cases in an initial filtering phase. In the refinement phase, a bounding-volume tree is applied to refine the detection results. Our experiments show that GPCD works well for the purpose of predictive collision detection.

#### **TA1-P3.23 ADAPTIVE VIDEO NEWS STORY TRACKING BASED ON EARTH MOVER'S DISTANCE**

*Mats Uddenfeldt, Uppsala University, Sweden; Keiichiro Hoashi, Kazunori Matsumoto, Fumiaki Sugaya, KDDI R&D Laboratories, Inc, Japan*

This paper proposes an adaptive system for video news story tracking based on the Earth Mover's Distance (EMD). When an interesting story appears in the news, it is flagged manually as a topic for tracking. Our system then tracks the events as they unfold over time and present accumulated results to the user for feedback. This feedback is used to adapt the topic model to changes in the tracked story. EMD provides the system with a robust way of performing many-to-many matching of news stories independent of the temporal order of their contents. This is particularly suitable in the news genre as stories often are subjected to video editing between shows. Experiments have been run with a range of topics and show promising results.

#### **TA1-P3.24 IMAGE CONTENT CLUSTERING AND SUMMARIZATION FOR PHOTO COLLECTIONS**

*Cheng-Hung Li, Chih-Yi Chiu, Chun-Rong Huang, Chu-Song Chen, Lee-Feng Chien, Academia Sinica, Taiwan*

Rapid growth of digital photography in recent years spurred the need of photo management tools. In this study, we propose an automatic organization framework for photo collections based on image content, so that a novel browsing experience is provided for users. For each photograph, human faces, together with corresponding clothes and nearby regions are located. We extract color histograms of these regions as the image content feature. Then a similarity matrix of a photo collection is generated according to temporal and content features of those photographs. We perform hierarchical clustering based on this matrix, and extract duplicate subjects of a cluster by introducing the contrast context histogram (CCH) technique. The experimental results show that the developed framework provides a promising result for photo management.

#### **TA1-P3.25 TRACKING OF HUMAN BODY JOINTS USING ANTHROPOMETRY**

*Alexei Gritai, Mubarak Shah, University of Central Florida, United States*

Most of the methods for human motion tracking are based on the modeling of human dynamics in action execution. In even small example space of human activities, the variation in action execution requires us to model a large number of uncertainties. This paper proposes a novel approach for motion tracking that avoids the tedious work of modeling human kinematics. This approach is based on the anthropometric and multi-view geometric constraints, successfully employed in the action recognition framework. The performance of this method is demonstrated on several different human actions.

#### **TA1-P3.26 A ROBUST METHOD FOR TV LOGO TRACKING IN VIDEO STREAMS**

*Jinqiao Wang, Chinese Academy of Sciences, China; Lingyu Duan, Institute for Infocomm Research, Singapore; Zhenglong Li, Jing Liu, Hanqing Lu, Chinese Academy of Sciences, China; Jesse S. Jin, University of Newcastle, Australia*

Most broadcast stations rely on TV logos to claim video content ownership or visually distinguish the broadcast from the interrupting commercial block. Detecting and tracking a TV logo is of interest to TV commercial skipping applications and logo-based broadcasting surveillance (abnormal signal is accompanied by logo absence). Pixel-wise difference computing within predetermined logo regions cannot address semi-transparent TV logos well for the blending effects of a logo itself and inconstant background images. Edge-based template matching is weak for semi-transparent ones when incomplete edges appear. In this paper we present a more robust approach to detect and track TV logos in video streams on the basis of multispectral images gradient. Instead of single frame based detection, our approach makes use of the temporal correlation of multiple consecutive frames. Since it is difficult to manually delineate logos of irregular shape, an adaptive threshold is applied to the gradient image in subpixel space to extract the logo mask. TV logo tracking is finally carried out by matching the masked region with a known template. An extensive comparison experiment has shown our proposed algorithm outperforms traditional methods such as frame difference, single frame-based edge matching. Our experimental dataset comes from part of TRECVID2005 news corpus and several Chinese TV channels with challenging TV logos.

#### **TA1-P3.27 AN AUTOMATIC CLASSIFICATION SYSTEM APPLIED IN MEDICAL IMAGES**

*Bo Qiu, ChangSheng Xu, Qi Tian, Institute for Infocomm Research, Singapore*

In this paper, a multi-class classification system is developed for medical images. We have mainly explored ways to use different image features, and compared two classifiers: Principle Component Analysis (PCA) and Supporting Vector Machines (SVM) with RBF (radial basis functions) kernels. Experimental results showed that SVM with a combination of the middle-level blob feature and low-level features (down-scaled images and their texture maps) achieved the highest recognition accuracy. Using the 9000 given training images from ImageCLEF05, our proposed method has achieved a recognition rate of 88.9% in a simulation experiment. And according to the evaluation result from the ImageCLEF05 organizer, our method has achieved a recognition rate of 82% over its 1000 testing images.

#### **TA1-P3.28 SPECIAL EFFECTS IN FILM/VIDEO MAKING: A NEW MEDIA INITIATIVE PROJECT**

*Chun-hao Wang, Yongjin Wang, Meifeng Lian, Bruce Elder, Ryerson University, Canada; Xiaou Tang, Microsoft Research Asia, ; Ling Guan, Ryerson University, Canada*

We present a system and a set of tools for producing special effects in film/video making by applying image processing and human centered computing techniques. A combination of shot detection, object segmentation, background generation, and image warping techniques are used. The user selects the image frame or the object of interest, and the image warp transformation to be used from the GUI. Wrapping can be performed either on a whole image sequence or on an object of interest in the sequence. In the latter, the object is first segmented and its motion tracked. Object segmentation is achieved either by snakes or graph cuts. Steerable Pyramid background generation is then used to fill in the portion cut from the foreground.

#### **TA1-P3.29 BROADCAST VIDEO PROGRAM SUMMARIZATION USING FACE TRACKS**

*Kadir Peker, Mitsubishi Electric Research Laboratories, United States; Isao Otsuka, Mitsubishi Electric Corporation, Japan; Ajay Divakaran, Mitsubishi Electric Research Laboratories, United States*

We present a novel video summarization and skimming technique using face detection on broadcast video programs. We take the faces in video as our primary target as they constitute the focus of most consumer video programs. We detect face tracks in video and define face-scene fragments based on start and end of face tracks. We define a fast-forward skimming method using frames selected from fragments, thus covering all the faces and their interactions in the video program. We also define novel constraints for a smooth and visually representative summary, and construct longer but smoother summaries.

#### **TA1-P3.30 ON THE USE OF TIME-FREQUENCY REPRESENTATION IN MULTICOMPONENT SIGNAL SEPARATION**

*Braham Barkat, Petroleum Institute, United Arab Emirates; Farook Sattar, Nanyang Technological University, Singapore; Karim Abed-Meraim, Telecom-Paris, France*

In this paper, we address the problem of separating unknown multicomponent signals from their instantaneous mixtures. Using linear time-frequency (TF) representation of the mixtures along with vectors classification scheme provide us a simple and efficient technique to separate multicomponent signals. The proposed algorithm can handle monocomponent as well as multicomponent sources and its assumptions about the mixing matrix are more relaxed compared to other existing TF based algorithms. The source separation results for the mixed synthetic signals as well as mixed real audio signals, such as mixture of speech and music, are shown to illustrate the validity and efficiency of the proposed scheme.

### **TA1-P4: Tuesday, July 11, 09:30 - 10:30**

#### **TA1-P4 Multimedia Computing and Software II (Poster)**

Time: Tuesday, July 11, 09:30 - 10:30

Place: Toronto III

Chair: Mark Drew, Simon Fraser University

#### **TA1-P4.31 TEMPLATE-BASED SEMI-AUTOMATIC PROFILING OF MULTIMEDIA APPLICATIONS**

*Christophe Poucet, IMEC, Belgium; David Atienza, Universidad Complutense Madrid - Dacya, Spain; Francky Catthoor, IMEC, Belgium*

Modern multimedia applications possess a very dynamic use of the memory hierarchy depending on the actual input, therefore requiring run-time profiling techniques to enable optimizations. Because they can contain hundreds of thousands of lines of complex object-oriented specifications, this constitutes a tedious time-consuming task since the addition of profilecode is usually performed manually. In this paper, we present a high-level library-based approach for profiling both statically and dynamically defined variables using templates in C++. Our results in the visual texture coder of the MPEG4 standard show that using the information it provides, we can easily achieve 70.56% energy savings and 19.22% memory access reduction.

#### **TA1-P4.32 COUNTING OF VIDEO CLIP REPETITIONS USING A MODIFIED BMH ALGORITHM: PRELIMINARY RESULTS**

*Silvio Guimaraes, Renata Coelho, Anne Torres, PUC Minas, Brazil*

In this work, we cope with the problem of identifying the number of repetitions of a specific video clip in a target video clip. Generally, the methods that deal with this problem can be subdivided into methods that use: (i) video signatures afterward the step of temporal video segmentation; and (ii) string matching algorithms afterward transformation of the video frame content into a feature values. Here, we propose a modification of the fastest exact string matching algorithm, the Boyer-Moore-Horspool, to count video clip repetitions. We also present some experiments to validate our approach, mainly if we are interested in found identical video clips according to spatial and temporal features.

**TA1-P4.33 MOVIE-BASED MULTIMEDIA MATRIX LIBRARY**

*Dmitry Vazhenin, Alexander Vazhenin, University of Aizu, Japan*

The paper describes a library supporting effective programming and design of matrix algorithms and programs. The important feature of proposed library is the visual algorithm representation using a movie-based approach. The user has a deal with special multimedia objects, each of which can generate an executable code as well as produce animation frames. These objects build an algorithmic skeleton representing the steps of computation. In this paper, we show the main features of movie-based programming as well as describe the movie-based matrix library. Examples of the library usage are also presented.

**TA1-P4.34 A PATTERN-SEARCH METHOD FOR H.264/AVC CAVLC DECODING**

*Shau-Yin Tseng, Tien-Wei Hsieh, Industrial Technology Research Institute, Taiwan*

In this paper, a new implementation method is proposed for Context-Adaptive Variable Length Coding (CAVLC) used in H.264 Baseline Profile. We analyze the correlation between bit patterns and 4x4 (or 2x2) blocks and have an idea of a pattern-search method before CAVLC decoding. If a pattern is matched in our look-up table, we can skip the standard CAVLD procedure and reconstruct a block directly. However, if there is not any pattern matched in the table, we have to reconstruct a block by CAVLD. Our look-up tables are built up according to our statistics and analysis. The experimental results show that the performance can be improved 10% compared with the standard CAVLD procedure.

**TA1-P4.35 A CONDITION-BASED INTRA PREDICTION ALGORITHM FOR H.264/AVC**

*Jia-Wei Chen, Chun-Hao Chang, Chien-Chang Lin, Yi-Huan Ou Yang, Jiun-In Guo, Jinn-Shyan Wang, National Chung Cheng University, Taiwan*

This paper proposes a condition-based algorithm for H.264/AVC 4x4 intra prediction. Exploiting high correlation existed in neighboring intra prediction modes, we propose the three conditions to skip the less possible candidates in doing Intra4x4 block mode decision. When compared to the 9 prediction modes in the full search algorithm, the proposed algorithm can complete a 4x4 intra prediction using 4.4 prediction modes operation in average. The simulation result shows that the proposed algorithm can reduce computational complexity up to 44% at the cost of less than 0.1dB PSNR loss in average.

**TA1-P5: Tuesday, July 11, 09:30 - 10:30****TA1-P5 Analysis and Synthesis of Speech and Audio I (Poster)**

Time: Tuesday, July 11, 09:30 - 10:30

Place: Toronto III

Chair: Frank Wallhoff, Technische Universität München

**TA1-P5.36 MINIMUM PHONEME ERROR BASED FILTER BANK ANALYSIS FOR SPEECH RECOGNITION**

*Hao Huang, Jie Zhu, Shanghai Jiao Tong University, China*

In this paper the optimal filter-bank design method based on the Minimum Phoneme Error (MPE) criteria is investigated. We use Gaussian type filter bank for optimization and various parameters of the filters such as gain, bandwidth and center frequency are trained aiming at maximize the MPE objective function to reduce word error. Preliminary experimental results on a large vocabulary continuous Mandarin speech recognition task given in this paper showed that cepstral coefficients derived from the optimized filter bank parameters result in a superior performance for word accuracy compared with the traditional triangle shaped filter bank. The filters consistent with the MPE criteria are also illustrated.



**TA1-P5.37 A TEMPO FEATURE VIA MODULATION SPECTRUM ANALYSIS AND ITS APPLICATION TO MUSIC EMOTION CLASSIFICATION**

*Yuan-Yuan Shi, Xuan Zhu, Hyoung-Gook Kim, Ki-Wan Eom, Samsung Advanced Institute of Technology, China*

This paper proposes a tempo feature extraction method based on the long-term modulation spectrum analysis. To transform the modulation spectrum to a condensed feature vector, the log-scale modulation frequency coefficients are introduced. This idea aims at averaging the modulation frequency energy via the constant-Q filter-banks. Further it is pointed out that the feature can be extracted directly from the perceptually compressed data of digital music archives. To verify the effectiveness of the feature and its utility to music applications, the feature vector is used in a music emotion classification system. The system consisting two layers of Adaboost classifiers. In the first layer the conventional timbre features are employed. Then by adding the tempo feature in the second layer, the classification precision is improved dramatically. By this way the discriminability of the classifier based on the given features can be exploited extremely. The system obtains high classification precision on a small corpus. It proves that the proposed feature is very effective and computationally efficient to characterize the tempo information of music.

**TA1-P5.38 MUSICAL SIGNAL TYPE DISCRIMINATION BASED ON LARGE OPEN FEATURE SETS**

*Björn Schuller, Frank Wallhoff, Dejan Arsic, Gerhard Rigoll, Technische Universität München, Germany*

Automatic discrimination of musical signal types as speech, singing, music or drumbeats within audio streams is of great importance e.g. for radio broadcast stream segmentation. Yet, feature sets are largely discussed. We therefore suggest a large open feature set approach starting with systematical generation of 7k hi-level features based on MPEG-7 Low-Level-Descriptors and further feature contours. A subsequent fast Gain Ratio reduction followed by wrapper-based Floating Search leads to a strong basis of relevant features. Next, features are added by alteration and combination within genetic search. For classification we use Support-Vector-Machines proven reliable for this task. Test-runs are carried out on a task-specific database and the public Columbia SMD database and show significant improvements for each step of the suggested novel concept.

**TA1-P5.39 EMOTIONAL SPEECH SYNTHESIS USING SUBSPACE CONSTRAINTS IN PROSODY**

*Shinya Mori, Tsuyoshi Moriyama, Shinji Ozawa, Keio University, Japan*

An efficient speech synthesis method that uses subspace constraint in prosody is proposed. Conventional unit selection methods concatenate speech segments stored in database, that require enormous number of waveforms in synthesizing various emotional expressions with arbitrary texts. The proposed method employs principal component analysis to reduce the dimensionality of prosodic components, that also allows us to generate new speech that are similar to training samples. The subspace constraint assures that the prosody of the synthesized speech including F0, power, and speech length hold their correlative relation that training samples of emotional speech have. We assume that the combination of the number of syllables and the accent type determines the correlative dynamics of prosody, for each of which we individually construct the subspace. The subspace is then linearly related to emotions by multiple regression analysis that are obtained by subjective evaluation for the training samples. Experimental results demonstrated that only 4 dimensions were sufficient for representing the prosodic changes due to emotion at over 90% of the total variance. Synthesized emotion were successfully recognized by the listeners of the synthesized speech.

**TA1-P5.40 A RANK BASED METRIC OF ANCHOR MODELS FOR SPEAKER VERIFICATION**

*Yingchun Yang, Min Yang, Zhaohui Wu, Zhejiang University, China*

In this paper, we present an improved method of anchor models for speaker verification. Anchor model is the method that represent a speaker by his relativity of a set of other speakers, called anchor speakers. It was firstly introduced for speaker indexing in large audio database. We suggest a rank based metric for the measurement of speaker character vectors in anchor model. Different from conventional metric methods which consider each anchor speaker equally and compare the log likelihood scores directly, in our method the relative order of anchor speakers is exploited to characterize target speaker. Experiments have been taken on the YOHO database, and the EER of our method is 13.29% lower than that of conventional metric. Also, our method is more robust against the mismatching between test set and anchor set.



## **TA1-P5.41 AUTOMATIC SPEAKER SEGMENTATION USING MULTIPLE FEATURES AND DISTANCE MEASURES: A COMPARISON OF THREE APPROACHES**

*Margarita Kotti, Aristotle University of Thessaloniki, Greece; Luís Gustavo P.M. Martins, INESC Porto, Portugal; Emmanouil Benetos, Aristotle University of Thessaloniki, Greece; Jaime Cardoso, INESC Porto, Portugal; Constantine Kotropoulos, Aristotle University of Thessaloniki, Greece*

This paper addresses the problem of unsupervised speaker change detection. Three systems based on the Bayesian Information Criterion (BIC) are tested. The first system investigates the AudioSpectrumCentroid and the AudioWaveformEnvelope features, implements a dynamic thresholding followed by a fusion scheme, and finally applies BIC. The second method is a real-time one that uses a metric-based approach employing the line spectral pairs and the BIC to validate a potential speaker change point. The third method consists of three modules. In the first module, a measure based on second-order statistics is used; in the second module, the Euclidean distance and the Hotelling statistic are applied; and in the third module, the BIC is utilized. The experiments are carried out on a dataset created by concatenating speakers from the TIMIT database, that is referred to as the TIMIT data set. A comparison between the performance of the three systems is made based on t-statistics.

## **TA2-L1: Tuesday, July 11, 10:50 - 12:10**

### **TA2-L1 Digital Watermarking, Data Hiding and Steganography II (Lecture)**

Time: Tuesday, July 11, 10:50 - 12:10

Place: Carmichael

Chair: Jean-Luc Dugelay, Institut EURECOM

10:50 - 11:10

### **TA2-L1.1 CAPTURING-RESISTANT AUDIO WATERMARKING BASED ON DISCRETE WAVELET TRANSFORM**

*Seungjae Lee, Sang-Kwang Lee, Young-Ho Seo, Electronics and Telecommunications Research Institute (ETRI), Republic of Korea; Chang D. Yoo, Korea Advanced Institute of Science and Technology, Republic of Korea*

In this paper, we propose a wavelet-based audio watermarking algorithm that is robust against capturing attack. With a commercial capturing tool, it is possible to capture various audio contents without noticeable degradation, and thus can potentially facilitate the illegal distribution of the audio content. By adjusting the mean value of the lowest subband coefficients of the discrete wavelet transform (DWT) of the audio, the proposed watermark can survive capturing attack including sampling rate conversion, random cropping and compression. By incorporating a simple human auditory model, the inaudibility of the watermark achieved, and the detection probability is improved based on the difference information of extracted values. This is confirmed by experimental results.

11:10 - 11:30

### **TA2-L1.2 SPREAD-SPECTRUM SUBSTITUTION WATERMARKING GAME**

*Jean-Philippe Boyer, LSS, France; Pierre Duhamel, LSS/CNRS, France; Jacques Blanc-Talon, CTA/GIP, France*

In the integrity checking context of multimedia contents, a malicious user aims at devising a forged content in order to fool a watermarker by making him use as a genuine content. By considering that the watermark acts as an integrity stamp, the false-alarm probability to recover the watermark signature in a forged content is the criterion of interest. We study and solve a game for this criterion between a watermarker and a falsifier which is allowed to perform a substitution attack, i.e. replace the watermarked signal by a non-watermarked content. As for the watermarker, we are concerned with additive spread-spectrum (SS) embedding. Signals are modeled by parallel colored gaussian processes. Due to the intractability of the false-alarm probability, we resort to Chernoff bound as an alternative cost. Our study confirms some common heuristics: the best attacker choice is to substitute the watermarked host signal using a signal which has very close statistics to the original host signal. The best watermarker strategy is to embed the watermark into the weakest frequency power components of the host signal. We finally consider the consequences of these results in terms of frequency embedding domain for an image SS watermarking scheme which has to be robust to compression. This reveals notable differences with informed scalar quantized-based schemes.

11:30 - 11:50

### **TA2-L1.3 PROTECTION OF 3D OBJECT VISUAL REPRESENTATIONS**

*Jihane Bennour, Jean-Luc Dugelay, Institut Eurecom, France*

In this paper, we describe a new framework for watermarking 3D objects via their contour information. Unlike most conventional 3D object watermarking techniques, for which both insertion and extraction of the mark are performed on the object itself (3D/3D approach), we propose an asymmetrical 3D/2D procedure. The goal of our work is to retrieve information (originally hidden in the apparent 3D silhouette of the object) from resulting images or videos having used the synthetic object, thus protecting the visual representations of the object. After developing theoretical and practical key-points of this 3D object watermarking scheme, we present the results of some preliminary experiments.

11:50 - 12:10

### **TA2-L1.4 IMAGE WATERMARKING BASED ON GENETIC ALGORITHM**

*Zhicheng Wei, Hao Li, Jufeng Dai, Sashuang Wang, Tianjin University, China*

In order to improve the robustness and imperceptibility of the image spread spectrum watermark algorithm, a new approach for optimization in 8x8 DCT domain using genetic algorithm (GA) was proposed. GA was used to choose the AC coefficients, which were modified to embed the spread spectrum watermark. The bands were varied to find the most suitable for image with different characteristics. Performance improvement with respect to existing algorithm is obtained by GA adaptive global search. The experimental results show that the proposed algorithm yields a watermark that is invisible to human eyes and robust to various image manipulation, and show that some special positions are the best choices for embedding the watermark. The authors also compare their experimental results with the results of the previous work of others.

## **TA2-L2: Tuesday, July 11, 10:50 - 12:10**

### **TA2-L2 Multi-User Collaboration and Cross Layer Optimization in Wireless Multimedia Communication II** (Special Session)

Time: Tuesday, July 11, 10:50 - 12:10

Place: Johnston

Co-Chairs: Zhu Li, Motorola Labs, Pascal Frossard, EPFL and Aggelos Katsaggelos, Northwestern University

10:50 - 11:10

### **TA2-L2.1 CROSS LAYER ADAPTATION FOR H.264 VIDEO MULTICASTING OVER WIRELESS LAN**

*Zhengye Liu, Polytechnic University, United States; Hang Liu, Thomson, Inc., United States; Yao Wang, Polytechnic University, United States*

This paper describes cross-layer optimization strategies and simulation results for H.264 video multicast over wireless LAN. The proposed scheme takes into account the varying channel conditions of multiple users, and dynamically allocates available bandwidth between source coding and channel coding. In particular, source coding parameters (intra update and quantization) and application-layer FEC code rate are chosen jointly to optimize a multicast performance criterion, based on feedback from all multicast receivers. Two performance criteria for video multicast are investigated and compared.

11:10 - 11:30

**TA2-L2.2 CROSS-LAYER CONGESTION CONTROL, SCHEDULING AND POWER CONTROL DESIGN IN MULTIHOP NETWORKS WITH RANDOM ACCESS**

*Chengnian Long, Hong Kong University of Science and Technology, Hong Kong SAR of China; Xinping Guan, Yanshan University, China; Bo Li, Hong Kong University of Science and Technology, Hong Kong SAR of China*

This paper considers optimal cross-layer rate control, scheduling design and power control for multi-hop wireless networks. The problem of optimal rate control, link scheduling, and link transmission power for all active time slot is formulated as a network utility maximization problem. In wireless multi-hop networks, the link capacity is a function of link scheduling and transmission power with time-varying and nonlinear properties. Those characteristic poses much challenge in joint design. To solve the non-convex and non-separable nonlinear program problem, a two time-scale distributed optimization approach is presented. By dual decomposition and gradient method, the NUM problem naturally decomposes into three subproblems: congestion control, scheduling design and power control. They interact through congestion price. The global convergence of this algorithm is proven. This paper presents a step towards a systematic approach to jointly design TCP congestion control algorithms, scheduling design and power control.

11:30 - 11:50

**TA2-L2.3 IMPLEMENTATION AND EVOLUTION OF PACKET STRIPING FOR MEDIA STREAMING OVER MULTIPLE BURST-LOSS CHANNELS**

*Gene Cheung, HP Laboratories, Japan; Puneet Sharma, Sung-Ju Lee, HP Labs Palo Alto, United States*

Modern mobile devices are multi-homed with WLAN and WWAN communication interfaces. In a community of nodes with such multi-homed devices --- locally inter-connected via high-speed WLAN but each globally connected to larger networks via low-speed WWAN, striping high-volume traffic from remote large networks over a bundle of low speed WWAN links can overcome the bandwidth mismatch problem between WLAN and WWAN. In our previous work, we showed that a packet striping system for such multi-homed devices --- a mapping of delay-sensitive packets by an intermediate gateway to multiple channels using combination of retransmissions (ARQ) and forward error corrections (FEC) --- can dramatically enhance the overall performance. In this paper, we improve upon a previous algorithm in two respects. First, by introducing two-tier dynamic programming tables to memoize computed solutions, packet striping decisions translate to simple table lookup operations given stationary network statistics. Doing so drastically reduces striping operation complexity. Second, new weighting functions are introduced into the hybrid ARQ/FEC algorithm to drive the long-term striping system evolution away from pathological local minima that are far from the global optimum. Results show the new algorithm performs efficiently and gives improved performance by avoiding local minima compared to the previous algorithm.

11:50 - 12:10

**TA2-L2.4 CROSS-LAYER DESIGN OF SOURCE RATE CONTROL AND QOS-AWARE CONGESTION CONTROL FOR WIRELESS VIDEO STREAMING**

*Peng Zhu, Tsinghua University, China; Wenjun Zeng, University of Missouri-Columbia, United States; Chunwen Li, Tsinghua University, China*

Cross-layer design has been used in streaming video over the wireless channels to optimize the overall system performance. In this paper we extend our previous work (i.e. joint design of source rate control and congestion control for video streaming over the Internet) in [1] and propose a cross-layer design approach for wireless video streaming. By jointly designing the source rate control at the application layer and congestion control at the transport layer, and taking advantage of MAC layer information, our approach can avoid the throughput degradation caused by transmission error of the wireless channel, and better support the QoS requirements of the application. Simulation results show that the proposed mechanism can significantly improve the playback quality of the application, while maintaining good performance of the transport protocol.

**TA2-L3 Multimedia in WWW (Lecture)**

Time: Tuesday, July 11, 10:50 - 12:10

Place: Thomson

Chair: Yong Rui, Microsoft

10:50 - 11:10

**TA2-L3.1 WEBDOVE: A WEB-BASED COLLABORATION SYSTEM FOR PHYSICAL TASKS**

*Weiyi Yang, Jiazhi Ou, Carnegie Mellon University, United States; Yong Rui, Microsoft Research, United States; Jie Yang, Carnegie Mellon University, United States*

While many systems are available for audio-visual people collaboration and data collaboration, systems for collaboration on physical objects are few. In this paper, we present WebDOVE, a system designed to address the needs of collaborative physical tasks. WebDOVE supports both live video streams and pen-based gesture recognition in multi-party bi-directional communication via inexpensive web cameras. WebDOVE allows distributed collaborators to draw over video streams to produce and interpret pointing and representational gestures as readily as they do in face-to-face settings. To accommodate potential diverse platform requirements from different participants, WebDOVE is designed to be a web-based platform-independent and browser-independent collaboration solution. We show via experiments that despite WebDOVE's platform independency, it requires moderate network bandwidth and CPU load, which make WebDOVE a practical solution for real-world applications.

11:10 - 11:30

**TA2-L3.2 PING: A GROUP-TO-INDIVIDUAL DISTRIBUTED MEETING SYSTEM**

*Yong Rui, Eric Rudolph, Li-Wei He, Rico Malvar, Michael Cohen, Ivan Tashev, Microsoft Research, United States*

Group-to-individual (G2I) distributed meeting is an important but understudied area. Because of the asymmetry between different parties in G2I meetings, it has two unique challenges: 1) the remote participant tends to be ignored by the local participants; and 2) the remote participant has inferior audio, video, and data experience than the local participants. To address these issues, in this paper we present PING, a system explicitly designed for G2I distributed meetings that combines recent advances in both hardware, e.g., microphone arrays, remote person stand-in devices, and software, e.g., audio-video processing, to improve users' G2I meeting experience. We report how PING addresses the above two challenges and its system design and implementation.

11:30 - 11:50

**TA2-L3.3 RESP: SHORTEST-PATH-BASED CACHE REPLACEMENT IN A TRANSCODING PROXY**

*Hao-Ping Hung, Ming-Syan Chen, National Taiwan University, Taiwan*

In this paper, we discuss the cache replacement policy in a multimedia transcoding proxy. Unlike the cache replacement for conventional web objects, to replace some elements with others in the cache of a transcoding proxy, we should further consider the inter-relationship among the cached items. To maintain the inter-relationship and to perform cache replacement, we propose in this paper the RESP framework (standing for Replacement with Shortest Path). The RESP framework contains two primary components, i.e., procedure MASP (standing for Minimum Aggregate Cost with Shortest Path) and algorithm EBR (standing for Exchange-Based Replacement). Procedure MASP maintains the inter-relationship using a shortest path table, whereas algorithm EBR performs cache replacement according to an exchanging strategy. The experimental results show that the RESP framework can approximate the optimal cache replacement with much lower execution time for processing user queries.

11:50 - 12:10

**TA2-L3.4 ROBUST DISTRIBUTED MULTI-POINT VIDEO CONFERENCING OVER ERROR-PRONE CHANNELS**

*Meng Chen, Guan-Ming Su, Min Wu, University of Maryland, College Park, United States*

In this paper, we propose a novel multi-point video conferencing system through error-prone channels, where the aggregation of multiple video streams and resource allocation are performed in a distributed manner. Video stream combiners, which are located in different geographical areas and serve as portals for conferees, aggregate incoming streams supplied by local users with other streams aggregated from nearby video stream combiners. A distributed multi-stream error protection scheme is performed in each video stream combiner to minimize the maximal expected video distortion among all aggregated streams. The simulation results demonstrate that our proposed scheme outperforms the traditional multicasting scheme by 1dB ~ 1.4dB in terms of average PSNR.

**TA2-L4 Audio and Speech Processing II (Lecture)**

Time: Tuesday, July 11, 10:50 - 12:10

Place: Toronto I

Chair: George Tzanetakis, University of Victoria

10:50 - 11:10

**TA2-L4.1 PERCEPTUALLY ENHANCED BIT-PLANE CODING FOR SCALABLE AUDIO**

*Rongshan Yu, Te Li, Susanto Rahardja, Institute for Infocomm Research, Singapore*

The MPEG-4 Scalable to Lossless (SLS) audio coding is recently being developed to provide a unified solution for high - compression perceptual audio coding and high-quality lossless audio coding. SLS provides efficient Fine Granular Scalable (FGS) coding from AAC core layer to lossless, and achieves reasonable perceptual quality at its scalable coding range using a sequential bit-plane scanning method, which minimizes the audio distortion according to the spectral shape of the core layer quantization errors. In this paper, it is shown that the perceptual quality performance of SLS at intermediate rates can be further improved by incorporating psychoacoustic model into the bit-plane coding process. In addition, it is also found that such an improvement can be achieved by slightly tweaking the original bit-plane coding process of SLS and hence preserving its nice features such as compatibility to lossless coding and low complexity.

11:10 - 11:30

**TA2-L4.2 SPEECH MODELING WITH MAGNITUDE-NORMALIZED COMPLEX SPECTRA AND ITS APPLICATION TO MULTISENSORY SPEECH ENHANCEMENT**

*Amarnag Subramanya, University of Washington, United States; Zhengyou Zhang, Zicheng Liu, Alex Acero, Microsoft Research, United States*

A good speech model is essential for speech enhancement, but it is very difficult to build because of huge intra- and extra-speaker variation. We present a new speech model for speech enhancement, which is based on statistical models of magnitude-normalized complex spectra of speech signals. Most popular speech enhancement techniques work in the spectrum space, but the large variation of speech strength, even from the same speaker, makes accurate speech modeling very difficult because the magnitude is correlated across all frequency bins. By performing magnitude normalization for each speech frame, we are able to get rid of the magnitude variation and to build a much better speech model with only a small number of Gaussian components. This new speech model is applied to speech enhancement for our previously developed microphone headsets that combine a conventional air microphone with a bone sensor. Much improved results have been obtained.

11:30 - 11:50

**TA2-L4.3 ROBUST SPEAKER RECOGNITION USING SNR-AWARE SUBSPACE-BASED ENHANCEMENT AND PROBABILISTIC SVMs**

*Jia-Ching Wang, Jhing-Fa Wang, Wai-He Kuok, Hsiao-Ping Lee, Chung-Hsien Yang, National Cheng Kung University, Taiwan*

In this paper, we present a robust text-independent speaker recognition system. The proposed system mainly includes an SNR-aware subspace-based enhancement technique and probabilistic support vector machines (SVMs). First, we construct a perceptual filterbank from psycho-acoustic model and incorporate it with the subspace-based enhancement approach. The prior SNR of each subband within the perceptual filterbank is taken to decide the estimator's gain to effectively suppress environmental background noises. Next, this study uses probabilistic SVMs to identify the speaker from the enhanced speech. The superiority of the proposed system has been demonstrated by twenty speaker recognition from AURORA-2 database with in-car noises.

11:50 - 12:10

**TA2-L4.4 ACOUSTICALLY-DRIVEN TALKING FACE SYNTHESIS USING DYNAMIC BAYESIAN NETWORKS**

*Jianxia Xue, Jonas Borgstrom, University of California, Los Angeles, United States; Jintao Jiang, Lynne Bernstein, House Ear Institute, United States; Abeer Alwan, University of California, Los Angeles, United States*

Dynamic Bayesian Networks (DBNs) have been widely studied in multi-modal speech recognition applications. Here, we introduce DBNs into an acoustically-driven talking face synthesis system. Three prototypes of DBNs, namely independent, coupled, and product HMMs were studied. Results showed that the DBN methods were more effective in this study than a multilinear regression baseline. Coupled and product HMMs performed similarly better than independent HMMs in terms of motion trajectory accuracy. Audio and visual speech asynchronies were represented differently for coupled HMMs versus product HMMs.

**TA2-L5: Tuesday, July 11, 10:50 - 12:10**

**TA2-L5 Multimedia Adaptation and Cross-Layer QoS for Wireless Networks (Lecture)**

Time: Tuesday, July 11, 10:50 - 12:10

Place: Toronto II

Chair: Jianfei Cai, Nanyang Technological University

10:50 - 11:10

**TA2-L5.1 IMPROVING THE PERFORMANCE OF TCP WIRELESS VIDEO STREAMING WITH A NOVEL PLAYBACK ADAPTATION ALGORITHM**

*Antonios Argyriou, Georgia Institute of Technology, United States*

In this paper we propose a playback adaptation algorithm for video streaming with TCP in wireless networks where both handoffs and random wireless errors are possible. Primary task of the algorithm is the estimation of the expected latency for the delivery of the lost packets by TCP. After this procedure is performed, the client adjusts the playback rate according to our algorithm so that both buffer underflows and overflows can be avoided. The proposed algorithm is non-intrusive to the protocol stack since it does not require any modifications to the baseline TCP protocol. To validate the basic principles behind our idea, we provide a set of comparative experimental results with other TCP-based streaming systems in terms of the rate of underflow events and the PSNR.

11:10 - 11:30

**TA2-L5.2 PROXY-BASED SNR SCALABLE ERROR TRACKING FOR REAL-TIME VIDEO TRANSMISSION OVER WIRELESS BROADBAND**

*Hui-Ya Li, Wei-Kuo Chiang, National Chung Cheng University, Taiwan*

In this paper, we propose a proxy-based SNR scalable error tracking framework for real-time video transmission where the server is wired connected to Internet and the client is connected to Internet through wireless broadband networks. We assume that all errors (packet losses) result from wireless links, and wired links are assumed to be error-free. The client sends back NACKs with the information about base layer lost packets to the proxy via a feedback channel. Once the NACK is received, the proxy uses the motion data and the side information received from the streaming server to perform error tracking. We compare our approach to the original proxy-based error tracking scheme without scalability support at the same bitrate and bit error rate. Experimental results show that the proposed method can effectively improve performance.

11:30 - 11:50

**TA2-L5.3 A CONTEXT-AWARE APPROACH FOR MULTIMEDIA PERFORMANCE OPTIMIZATION USING NEURAL NETWORKS IN WIRELESS LAN ENVIRONMENTS**

*Po-Chiang Lin, Chiapin Wang, Tsung-Nan Lin, National Taiwan University, Taiwan*

Packet size is one of the most important factors that would affect the user-perceived multimedia QoS in the wireless LAN environments. The time-varying channel characteristics make it difficult to find the exact relationship between the packet size and the throughput and decide an optimal packet size in advance. Furthermore, every node would suffer different channel conditions. In this paper, we tackle this problem by an optimization approach. A context-aware framework is designed to optimize the packet size adaptively in order to maximize the throughput. In this approach each node abstracts its specific context via the throughput from the time-varying wireless environments. The obtained throughput information is the instantaneous integrated effect of all contexts in wireless LAN environments. This approach adopts neural networks to learn the complex nonlinear function between the packet size and the throughput and adaptively adjusts the packet size. Simulation results show that our method can cope with the time-varying wireless channel conditions and improve the perceived QoS of wireless multimedia services.

11:50 - 12:10

**TA2-L5.4 BANDWIDTH ESTIMATION IN WIRELESS LANS FOR MULTIMEDIA STREAMING SERVICES**

*Heung Ki Lee, Varrian Hall, Texas A&M University, United States; Ki Hwan Yum, University of Texas, San Antonio, United States; Kyoung Ill Kim, Electronics and Telecommunications Research Institute (ETRI), Republic of Korea; Eun Jung Kim, Texas A&M University, United States*

The popularity of multimedia streaming services via wireless networks presents major challenges in the management of network bandwidth. One challenge is to quickly and precisely estimate the available bandwidth for the varying streaming rates of layered and scalable multimedia services. Previous works based on wired networks are too burdensome to be applied in wireless networks. In this paper, a new method, IdleGap, is suggested to estimate the available bandwidth of a wireless LAN based on the information from a low layer in the protocol stack. We use a network simulation tool, NS-2, to evaluate our new method with various range of cross traffic and observation times. Our simulation results show that IdleGap accurately estimates the available bandwidth for all ranges of cross traffic (100Kbps ~ 1Mbps) with a very short observation time (10 seconds).

**TA2-L6: Tuesday, July 11, 10:50 - 12:10**

**TA2-L6 Media Signal Processing III (Lecture)**

Time: Tuesday, July 11, 10:50 - 12:10

Place: Varley

Chair: Oscar Au, Hong Kong University of Science & Technology

10:50 - 11:10

**TA2-L6.1 DEMOSAICKING USING VECTOR SPECTRAL MODEL**

*Rastislav Lukac, Konstantinos N. Plataniotis, University of Toronto, Canada*

A new edge-sensing demosaicking solution based on a novel vector spectral model is introduced in this paper. Using the vector spectral model, the proposed solution preserves the magnitude and the directional characteristics of the single-sensor captured image in both smooth and edge areas. Experimentation reported in this paper indicates that the proposed demosaicking solution produces visually pleasing color images and can outperform the powerful demosaicking schemes in terms of the performance.



11:10 - 11:30

**TA2-L6.2 A SINGLE HEISENBERG-GABOR BASED FIGURE-OF-MERIT BASED ON THE MODULATION TRANSFER FUNCTION OF DIGITAL IMAGING SYSTEMS**

*Corey Manders, Steve Mann, University of Toronto, Canada*

We propose a single figure-of-merit measure of resolution of a digital imaging system based on the work of Gabor in communication theory. Gabor's work was largely inspired by Heisenberg's developments in quantum theory. Gabor's results look simultaneously at the frequency and spatial domain of a signal, making it ideal for the measure of the modulation transfer function and point-spread function of an imaging system. As opposed to the crude "megapixel" measure which is bantered about in the marketplace, we suggest our figure-of-merit which more accurately represents the resolution of the system. Given that the resolution measure we propose is condensed into a single number rather than a function such as the modulation transfer function or the point spread function, it is our intent to propose this scientific evaluation as a means for typical consumers to fairly judge the resolution of a camera. Finally, we use this measure to compare common digital SLR cameras with varying lenses.

11:30 - 11:50

**TA2-L6.3 RESTORATION OF MOTION BLURRED IMAGES**

*Juwei Lu, Eunice Poon, EPSON Edge, EPSON Canada Limited, Canada; Konstantinos N. Plataniotis, University of Toronto, Canada*

In this paper, we present several algorithms developed for restoration of motion blurred images. We begin with a single-image based deblurring approach in the case of linear constant motion. This approach is a wavelet-based method with a novel  $L_p$ -norm regularization term. Due to the introduction of the wavelet and regularization techniques, the approach is rather robust against noise amplification during deconvolution. Then, we further developed a general multi-image based deblurring framework improved from recent works of [1]. The proposed framework is able to effectively take advantage of information contained in the multiple input images, even when they are blurred in the same direction - a case hard to be dealt with by traditional solutions. The proposed methods are evaluated on both simulated and real data, and the obtained experimental results indicate promising results.

11:50 - 12:10

**TA2-L6.4 IMAGE VECTOR QUANTIZATION INDICES RECOVERY USING LAGRANGE INTERPOLATION**

*Yung-Gi Wu, Chia-Hao Wu, Leader University, Taiwan*

Vector quantization (VQ) is an efficient coding algorithm due to its fast decoding efficiency. Indices of VQ will be lost during the transmission because of the signal interference. In this paper, we propose an efficient estimation method by using the Lagrange interpolation formula to recover the lost indices in image vector quantization codec. If the image or video has the limitation of the period of validity, re-transmitting the data wastes of time. Therefore, using the received correct data to estimate and recover the lost data is efficient in time-constraint situation such as network conference. For nature image or video, the pixels with its neighbors are correlative. Since the VQ partitions the image into sub-blocks and quantize them to form the indices to transmit, the correlation between adjacent indices is very strong. There are two important parts of the proposed method. One is preprocessing process and the other is the estimation process. In preprocessing, we modify the order of code-vectors in the VQ codebook to increase the correlation between neighboring vectors. On the second part, the recovery process on the decoder, using the Lagrange interpolation formula to constitute a polynomial to describe the tendency of VQ indices, and use the polynomial to estimate the lost VQ indices. The simulation results demonstrate that our method can efficiently estimate the lost indices in acceptable visual quality.

**TA2-P1 Multimedia Networking (Poster)**

Time: Tuesday, July 11, 10:50 - 11:50

Place: Toronto III

Chair: Ivan Lee, Ryerson University

**TA2-P1.1 A DECENTRALIZED KEY MANAGEMENT SCHEME IN OVERLAY MULTICAST NETWORK**

*Xingfeng Guo, Xiaodong Liu, Qionghai Dai, Tsinghua University, China*

The recent growth of the World Wide Web has sparked new research into using the Internet for novel types of group communication, like multiparty videoconferencing and real-time streaming. Multicast has the potential to be very useful, but it suffers from many problems like security. To achieve secure multicast communications, key management is one of the most critical problems. So far, most of multicast key management schemes have been proposed and most of them are centralized, which have the problem of “one point failure” and that the group controller is the bottleneck of the group. In order to solve these two problems, we propose a Decentralized Key Management Scheme (DKMS), using RSA key system as auxiliary keys. We analyze this scheme and find it has appropriate performance in security and scalability.

**TA2-P1.2 SIKAS: A SCALABLE DISTRIBUTED KEY MANAGEMENT SCHEME FOR DYNAMIC COLLABORATIVE GROUPS**

*Jiang Zhang, Jian-Guang Luo, Bin Li, Shi-Qiang Yang, Tsinghua University, China*

Some distributed collaborative key management protocols have been proposed to provide group communication privacy and data confidentiality for collaborative groups. However, most of them rekey on each member change, and the costs of group rekeying can be quite substantial for large groups with frequent membership changes. In this paper, we propose a scalable distributed key management scheme using a distributed one-way function tree named SIKAS which can significantly reduce the computation and communication costs of maintaining the group key based upon period-based group rekeying. A comparison with previous work has shown that SIKAS provides scalability and rekeying efficiency while preserving both distributed and collaborative properties.

**TA2-P1.3 DEGREE PRE-RESERVED HIERARCHICAL TREE FOR MULTIMEDIA MULTICAST**

*Nan Zhang, Yuan-Chun Shi, Bin Chang, Tsinghua University, China*

Overlay multicast tree is widely used to support large-scale real-time multimedia applications. The scalability and the robustness are two key issues in the overlay structure design. In this paper, we propose a Degree pre-reserved hierarchical tree for multimedia multicast, called DTree. It organizes the overlay tree in two hierarchies, and combines application-level multicast with IP multicast to achieve a low delay data delivery. A Degree pre-reserved mechanism is designed in DTree, which greatly shortens the time to resume the data flow, and achieve a fast recovery. Our simulation results show that DTree has a better delivery performance, compared with other overlay tree. It is quite responsive to the changes in the tree, and almost 3 times faster than other recovery strategies in some cases.

**TA2-P1.4 MOBILITY MANAGEMENT FOR UNTETHERED IMMERSIVE COMMUNICATIONS**

*Mehran Dowlatshahi, Farzad Safaei, Smart Internet Technology CRC / University of Wollongong, Australia*

In this paper we propose a system design for delivery of immersive communications to mobile wireless devices based on a distributed proxy model. It is demonstrated that this architecture addresses key technical challenges for the delivery of these services, that is, constraints on link capacity and power consumption in mobile devices. However, additional complexity is introduced with respect to mobility management. The paper proposes three possible methods for updating proxy assignments in response to mobility and compares the performance of these methods.

## **TA2-P1.5 EFFICIENT WIRELESS MULTICAST PROTOCOL WITH ORTHOGONAL CTS MODULATION SUPPORTING VIDEO CONFERENCING**

*Ju Wang, Hongsik Choi, Esther Hughes, Virginia Commonwealth University, United States; Yong Tang, University of Florida, United States*

Multicast based video conferencing in wireless networks is recently receiving more and more attention. In this paper, we propose a novel linker-level protocol, Multicast with Orthogonal CTS (MOCTS) protocol, to provide efficient wireless multicast. Our approach uses orthogonal Walsh code to modulate CTS packets at different receiving nodes. The modulated CTSs are simultaneously transmitted and decoded at the source node, thus considerably reducing multicast CTS collisions. We discussed the protocol design to resolve code collisions and a collision salvage technique which tolerates a certain degree of collisions. Simulation results show that 16 or more codes should be used, and collision salvage could be a very effective. The MOCTS protocol is further extended to multi-hop networks with a localgain- maximizing (DLGM) algorithm to control the propagation of multicast packets. Simulation results show that, with DLGM employed in MOCTS protocol, a 20% or more saving is possible in total transmission.

## **TA2-P1.6 INTERNET TRAFFIC CLASSIFICATION FOR SCALABLE QOS PROVISION**

*Junghun Park, University of Southern California, United States; Hsiao-Rong Tyan, Chung Yuan Christian University, Taiwan; C.-C. Jay Kuo, University of Southern California, United States*

A new scheme that classifies the Internet traffic according to their application types for scalable QoS provision is proposed in this work. The traditional port-based classification method does not yield satisfactory performance, since the same port can be shared by multiple applications. Furthermore, asymmetric routing and errors of modern measurement tools such as PCF and NetFlow degrades the classification performance. To address these issues, the proposed classification process consists of two steps: feature selection and classification. Candidate features that can be obtained easily by ISP are examined. Then, we perform feature reduction so as to balance the performance and complexity. As to classification, the REPTree and the bagging schemes are adopted and compared. It is demonstrated by simulations with real data that the proposed classification scheme outperforms existing techniques.

## **TA2-P1.7 MULTICAST OF REAL TIME MULTI-VIEW VIDEO**

*Li Zuo, Xi'an Jiaotong University, China; Jian-Guang Lou, Hua Cai, Jiang Li, Microsoft Research Asia, China*

As a recently emerging service, multi-view video provides a new viewing experience with high degree of freedom. However, due to the huge data amounts transferred, multi-view video's delivery remains a daunting challenge. In this paper, we propose a multi-view, video-streaming system based on IP multicasting. It can support a large number of users while still keeping a high degree of interactivity and low bandwidth consumption. Based on a careful user study, we have developed two schemes: one is for automatic delivery and the other for on-demand delivery. In automatic delivery, a server periodically multicasts special effect snapshots at a certain time interval. In on-demand delivery, a server delivers the snapshots based on distribution of users' requests. We conducted extensive experiments and user-experience studies to evaluate the proposed system's performance, and found the system can provide satisfying multi-view video service for users on a large scale.

## **TA2-P1.8 PACKET LOSS MODELING FOR PERCEPTUALLY OPTIMIZED 3D TRANSMISSION**

*Irene Cheng, Lihang Ying, Anup Basu, University of Alberta, Canada*

Transmissions over wireless and other unreliable networks can lead to packet loss. An area that has received limited research attention is how to tailor multimedia information taking into account the way packets are lost. We provide a brief overview of our research on designing a 3D perceptual quality metric integrating two important factors, resolution of texture and resolution of mesh, which control transmission bandwidth. We then suggest alternative strategies for packet 3D transmission of both texture and mesh. These strategies are then compared with respect to preserving 3D perceptual quality under packet loss in ad hoc wireless networks. Experiments are conducted to study how the time between consecutive packet transmission and packet size affects loss in wireless channels. A preliminary model for estimating the optimal packet size is then proposed.

**TA2-P1.9 A NEURAL NETWORK BASED ADAPTIVE ALGORITHM FOR MULTIMEDIA QUALITY FAIRNESS IN WLAN ENVIRONMENTS**

*Chiapin Wang, Tsung-Nan Lin, National Taiwan University, Taiwan*

This paper investigates multimedia quality fairness in a wireless LAN environment where channel are error-prone due to mobility and fading. The experimental results show that using fixed MAC arguments for nodes in heterogeneous channel conditions leads to unequal throughput performance and that may incur the degradation of multimedia QoS. To overcome the unfairness problem for provisioning QoS, we propose a cross-layer adaptation scheme by on-line adapting the multidimensional MAC-layer backoff parameters depending on the application-layer QoS requirements and the PHY-layer channel conditions. Our solution is based on an optimization approach which utilizes neural networks to learn the cross-layer function. Simulations results demonstrate that our adaptive scheme can tackle a variety of channel condition to provide fair throughput for nodes in heterogeneous channel conditions.

**TA2-P1.10 AN ANALYTICAL STUDY OF STREAM TAPPING PROTOCOLS**

*Jehan-François Pâris, University of Houston, United States; Darrell Long, University of California, Santa Cruz, United States*

We present the first analytic study of stream tapping protocols, a family of protocols that provide the most efficient way to distribute videos on demand at low to medium request arrival rates, say, less than ten requests per hour for a two-hour video. The main results of this study are analytical solutions for the optimal operational points of stream tapping, stream tapping with small client buffers, stream tapping with partial preloading and stream tapping with proactive streams. In addition we introduce a new stream tapping protocol with batching that caps the bandwidth requirements of stream tapping at high to very high arrival rates.

**TA2-P2: Tuesday, July 11, 10:50 - 11:50**

**TA2-P2 Multimedia Systems, Architectures and Hardware (Poster)**

Time: Tuesday, July 11, 10:50 - 11:50

Place: Toronto III

Chair: Aroutchelvame Mayilavelane, Ryerson University

**TA2-P2.11 HIGH PERFORMANCE FRACTIONAL MOTION ESTIMATION AND MODE DECISION FOR H.264/AVC**

*Chao-Yang Kao, Huang-Chih Kuo, Youn-Long Lin, National Tsing Hua University, Taiwan*

We propose a high performance architecture for fractional motion estimation and Lagrange mode decision in H.264/AVC. Instead of time-consuming fractional-pixel interpolation and secondary search, our fractional motion estimator employs a mathematical model to estimate SADs at quarter-pixel position. Both computation time and memory access requirements are greatly reduced without significant quality degradation. We propose a novel cost function for mode decision that leads to much better performance than traditional low complexity method. Synthesized into a TSMC 0.13 $\mu$ m CMOS technology, our design takes 56k gates at 100MHz and is sufficient to process 3200x2400 video sequences at 30 fps. Compared with a state-of-the-art design operating under the same frequency, ours is 30% smaller and has 18 times more throughput at the expense of only 0.05db in PSNR difference.

## **TA2-P2.12 AN EVENT-DRIVEN SPORTS VIDEO ADAPTATION FOR THE MPEG-21 DIA FRAMEWORK**

*Min Xu, Jiaming Li, Yiqun Hu, Liang-Tien Chia, Bu-Sung Lee, Deepu Rajan, Jianfei Cai, Nanyang Technological University, Singapore*

We present an event-driven video adaptation system in this paper. Events are detected by audio/video analysis and annotated by the description schemes (DSs) provided by MPEG-7 Multimedia Description Schemes (MDSs). And then, adaptation take account of users' preference of events and network characteristic to adapt video by event selection and frame dropping as following three steps: 1) the event information is parsed from MPEG-7 annotation XML file together with bitstream to generate generic Bitstream Syntax Description (gBSD). 2) Users' preference, Network Characteristic and Adaptation QoS (AQoS) are considered for making adaptation decision. 3) adaptation engine automatically parses adaptation decisions and gBSD to achieve adaptation. Different from most existing adaptation work, the system adapts video by interesting events according to users' preference. To achieve a generic adaptation solution, the system is developed following MPEG-7 and MPEG-21 standards. gBSD based adaptation avoids complex video computation. 30 students from various departments test the system with satisfaction. Although, the system is tested on basketball video adaptation so far, it is easy to extend to other video domains.

## **TA2-P2.13 AN IMAGE PROCESSING PIPELINE WITH DIGITAL COMPENSATION OF LOW COST OPTICS FOR MOBILE TELEPHONY**

*Nikolaos Bellas, Motorola, United States; Arnold Yanof, Freescale, United States*

As digital imaging becomes more prevalent in consumer products, the industry strives to reduce the cost and the complexity of imaging solutions and, at the same time, to improve the color quality and resolution of the images. The proliferation of imaging in mobile phones, digital cameras, webcams, toys, etc. creates the need for a low cost, small footprint image sensor. Nowadays, sensor optics account for approximately half the cost of an image sensor system for a camera phone, and their cost does not scale down as fast as the cost of semiconductors. In this paper, we describe the algorithms and the hardware implementation of a novel color processing chain that uses image processing techniques to compensate for the spatial variations in image attributes and quality due to low cost optics. If left uncorrected, these variations produce undesirable visual effects and lead to unacceptable image quality. Besides the correction of artifacts due to lenses, the image processing chain performs a sequence of corrections for real time dead pixel replacement, color correction, filtering, color space transformations for subsequent compression, etc.

## **TA2-P2.14 COMPLEXITY ANALYSIS OF H.264 DECODER FOR FPGA DESIGN**

*Tuomas Lindroth, Nastoooh Avessta, Jukka Teuhola, Tiberiu Seceleanu, University of Turku, Finland*

A major challenge in the design of any real time system is the proper selection of implementation and platform alternatives. In this paper, a suitable FPGA-based design of the H.264 decoder is presented. Since H.264 standard only specifies the syntax and semantics of the video stream and not the video codec itself, the selection process may be directed based upon the temporal complexity of different parts of the decoder. Here, we present the process flow of these parts using basic algebraic operators. The analysis of the required logic elements to implement the decoder, on various platforms, is presented.

## **TA2-P2.15 APPROXIMATING OPTIMAL VISUAL SENSOR PLACEMENT**

*Eva Hörster, Rainer Lienhart, University of Augsburg, Germany*

Many novel multimedia applications use visual sensor arrays. In this paper we address the problem of optimally placing multiple visual sensors in a given space. Our linear programming approach determines the minimum number of cameras needed to cover the space completely at a given sampling frequency. Simultaneously it determines the optimal positions and poses of the visual sensors. We also show how to account for visual sensors with different properties and costs if more than one kind is available, and report performance results.

## **TA2-P2.16 PERFORMANCE-COMPLEXITY ANALYSIS OF HIGH RESOLUTION VIDEO ENCODER AND ITS MEMORY ORGANIZATION FOR DSP IMPLEMENTATION**

*Zhigang Yang, Harbin Institute of Technology, China; Wen Gao, Chinese Academy of Sciences, China; Yan Liu, Harbin Institute of Technology, China*

This paper first analyses the relationship between performance and complexity of several state-of-the-art coding algorithms for high resolution videos. Based on the coding efficiency comparison under different config parameters and the intra mode usage in P/B frame, this paper presents a practical scheme to improve the coding speed with slight quality loss. And a DSP-oriented two-level internal memory organization is also proposed to keep pipeline processing. In such organization, block correlation caused by motion vector predictions is lightened while keeping almost the same performance as the original.

**TA2-P2.17 AUTOMATIC COUNTING OF INTERACTING PEOPLE BY USING A SINGLE UNCALIBRATED CAMERA**

*Senem Velipasalar, Princeton University, United States; Ying-Li Tian, Arun Hampapur, IBM, United States*

Automatic counting of people, entering or exiting a region of interest, is very important for both business and security applications. This paper introduces an automatic and robust people counting system which can count multiple people who interact in the region of interest, by using only one camera. Two-level hierarchical tracking is employed. For cases not involving merges or splits, a fast blob tracking method is used. In order to deal with interactions among people in a more thorough and reliable way, the system uses the mean shift tracking algorithm. Using the first-level blob tracker in general, and employing the mean shift tracking only in the case of merges and splits saves power and makes the system computationally efficient. The system setup parameter can be automatically learned in a new environment from a 3 to 5 minute-video with people going in or out of the target region one at a time. With a 2GHz Pentium machine, the system runs at about 33fps on 320x240 images without code optimization. Average accuracy rates of 98.5% and 95% are achieved on videos with normal traffic flow and videos with many cases of merges and splits, respectively.

**TA2-P2.18 DESIGN OF AUDIO/VIDEO DECODER FOR THE T-DMB RECEIVER**

*Bontae Koo, Nakwoong Eum, Sekho Lee, Juhyun Lee, Jinkyu Kim, Minseok Choi, Kihyuk Park, Seongmin Kim, Duckwhan Kim, Electronics and Telecommunications Research Institute (ETRI), Republic of Korea*

We present a low-power architectural MPEG-4 part-10 AVC/H.264 video and MPEG-4 BSAC audio decoder chip capable of delivering high-quality and high-compression in wireless multimedia applications such as DMB (digital multimedia broadcasting). This AV decoder chip comprise all units required for T-DMB multimedia decoding such as system demultiplexer, MPEG-4 AVC/H.264 video decoder and MPEG-4 BSAC audio decoder. The proposed Audio/Video decoder has low power consumption and has been implemented using a standard-cell library in 0.18um 1P6M CMOS technology.

**TA2-P2.19 AN ARCHITECTURE DESIGN OF THRESHOLD-BASED BEST-BASIS ALGORITHM**

*Arouthelvame Mayilavelane, Kaamran Raahemifar, Ryerson University, Canada*

The best-basis algorithm has gained much importance on textured-based image compression and denoising of signals. In this paper, an architecture for the wavelet-packet based best-basis algorithm for images is proposed. The paper also describes the architecture for best-tree selection from 2D wavelet packet decomposition. The precision analysis of the proposed architecture is also discussed and the result shows that increase in the precision of input pixel greatly increases the Signal-to-Noise Ratio (SNR) per pixel whereas increase in the precision of filter coefficient does not greatly help in improving the SNR value. The proposed architecture is described in VHDL at the RTL level, simulated successfully for its functional correctness and implemented in an FPGA.

**TA2-P2.20 DATA HIDING FOR SPEECH BANDWIDTH EXTENSION AND ITS HARDWARE IMPLEMENTATION**

*Fan Wu, Siyue Chen, Henry Leung, University of Calgary, Canada*

Most of the current speech transmission systems are only able to deliver speech signals in a narrow frequency band. This narrowband speech is characterized by a thin and muffled sound. In this paper, we propose a data hiding scheme to artificially extend the speech bandwidth so that the speech quality can be improved. The hardware perspectives of implementing the proposed scheme are also discussed. A Xilinx MicroBlaze soft processor is used in this study. To balance the trade off between logic resource consumption and processing speed, the data hiding scheme are mainly implemented in application software while fast Fourier transform (FFT) is designed into a hardware acceleration model. Experimental results demonstrate the effectiveness of the proposed data hiding scheme.

**TA2-P3 Multimedia Applications II (Poster)**

Time: Tuesday, July 11, 10:50 - 11:50

Place: Toronto III

Chair: Qusay Mahmoud, University of Guelph

**TA2-P3.21 ONLINE TRAINING-ORIENTED VIDEO SHOOTING NAVIGATION SYSTEM BASED ON REAL-TIME CAMERAWORK EVALUATION**

*Masahito Kumano, Ryukoku University, Japan; Kuniaki Uehara, Yasuo Arik, Kobe University, Japan*

In this paper, we propose an online training-oriented video shooting navigation system focused on camerawork based on video grammar by real-time camerawork evaluation to train users shooting nice shots for the later editing work. In this system, the processing speed must be very high so that we use a luminance projection correlation and a structure tensor method to extract the camerawork parameters in real-time. From the results of camerawork analysis, the results of each frame are classified into 7 camerawork types and the system issues 6 types of alarms and navigates users along the specified shot depending on camerawork based on video grammar in real-time while shooting the shot. Thereby, users can naturally acquire shooting style by trying to decrease alarms of improper camerawork without a consideration of the video grammar.

**TA2-P3.22 AUTOMATIC ADDRESSEE IDENTIFICATION BASED ON PARTICIPANTS' HEAD ORIENTATION AND UTTERANCES FOR MULTIPARTY CONVERSATIONS**

*Yoshinao Takemae, NTT Corporation, Japan; Shinji Ozawa, Keio University, Japan*

We propose a method that uses the participants' head orientation and utterances for automatically identifying the addressee of each utterance in face-to-face multiparty conversations, such as meetings. First, each participant's head orientation is determined through vision-based detection and the presence/absence of utterances is extracted using the power of voices captured by microphones. Second, gaze direction (whom each participant is looking at) is estimated from just detected head orientation using the Support Vector Machine. Third, several related features such as amount and frequency of gaze and eye contact are calculated in each utterance interval. Finally, a Bayesian Network is used to classify each utterance into one of two types of utterances: (a) the speaker is addressing a single participant and (b) the speaker is addressing all participants. Experiments on addressee estimation with 3-person conversations confirm the usefulness of our method.

**TA2-P3.23 MEDIA SYNCHRONIZATION METHOD FOR VIDEO HYPERMEDIA APPLICATION BASED ON EXTENDED EVENT MODEL**

*Hironobu Abe, Mitsubishi Electric Corporation, Japan; Hiroshi Shigeno, Ken-ichi Okada, Keio University, Japan*

This paper describes a proposal of an extended event model using a media synchronization method for video hypermedia applications. In this extended event model, video and metadata are synchronized by periodically inserting event information in the video multiplex. We considered the following design policies: 1) a model that is independent of the video format and delivery method, 2) the synchronization accuracy can be tuned depending on the purpose and use of the metadata. We designed the extended event model based on the above design policies, and implemented this model as an encode/decode library for Windows Media. Based on this model we developed a video hypermedia system prototype and performed evaluation experiments. The evaluation results of real time synchronization performance of the system prototype showed that in the case of sports video content a synchronization accuracy of 100 msec between video and metadata, makes our method effective for use in video hypermedia applications.

**TA2-P3.24 A DECISION MECHANISM FOR PROCESSING MULTIMODAL SERVICES IN FUTURE GENERATION NETWORK**

*Yang Li, H. Anthony Chan, University of Cape Town, South Africa*

Communication technologies, old or new, are pushing the development of telecommunication industry. Thus, technically performing a multimodal service session (eg. one end is involved with data while the other end is with multimedia) is no longer a problem. People are gaining interests in managing these multimodal services by considering the choice and preference of users. This research brings in a four-option decision mechanism to intelligently process a multimodal service and may ensure a successful and friendly communication session for two reasons. Firstly, this mechanism provides four extra options for a service that may fail in current communication systems. After three extra tries, the service session will most probably succeed. Secondly, the user make decisions themselves by setting the rules of how to make proper decisions in advance.



## **TA2-P3.25 CLASSROOM MULTIMEDIA INTEGRATION FOR ADVANCED E-PRESENTATIONS**

*Weihong Li, City University of New York - Graduate Center, United States; Hao Tang, Chad McKittrick, Zhigang Zhu, City University of New York - City College, United States*

We have developed several basic components for our Virtualized Classroom project: automated data collection, intelligent media integration, and flexible user interfaces. As an example of media integration, we show that when the PPT slides are projected onto the digital whiteboard where the instructor annotates, modifies, or expands the PPT presentation, geometrically registering high-resolution images from the two sources yields high quality digital presentations (e-presentations). We use a low-cost digital camera as a bridge to align the two sources and present a hybrid registration approach to align the PPT and handwriting images. Experimental results are presented to validate our approach.

## **TA2-P3.26 TRICODES: A BARCODE-LIKE FIDUCIAL DESIGN FOR AUGMENTED REALITY MEDIA**

*Jonathan Mooser, Suyu You, Ulrich Neumann, University of Southern California, United States*

Visual markers, or fiducials, have become one of the most common methods of camera pose estimation in Augmented Reality (AR) media. Many present day fiducial-based AR systems use arbitrary patterns, such as simple line drawings or alpha-numeric characters, and require that an application be "trained" to recognize its pattern set. These techniques work well on a small scale, but as the number of fiducials grows, accuracy and performance degrade. We describe a new fiducial design called TriCodes that, like a barcode, provides a systematic way of printing and identifying a vast library of patterns. We compare TriCodes to the popular ARToolkit package, demonstrating its advantages in the presence of large numbers of fiducials.

## **TA2-P3.27 VIDEO AND AUDIO EDITING FOR MOBILE APPLICATIONS**

*Ari Hourunranta, Asad Islam, Fehmi Chebil, Nokia, Finland*

Video content creation and consumption have been increasingly available for the masses with the emergence of handheld devices capable of shooting, downloading, and playing videos. Video editing is a natural and necessary operation that is most commonly employed by users for finalizing and organizing their video content. With the constraints in processing power and memory, conventional spatial domain video editing is not a solution for mobile applications. In this paper, we present a complete video editing system for efficiently editing video content on mobile phones using compressed domain editing algorithms. A critical factor from usability point of view is the processing speed of the editing application. We show that with the proposed compressed domain editing system, typical video editing operations can be performed much faster than real-time on today's S60 phones.

## **TA2-P3.28 ON TRAINING NEURAL NETWORK ALGORITHMS FOR ODOR IDENTIFICATION FOR FUTURE MULTIMEDIA COMMUNICATION SYSTEMS**

*Ki-Hyeon Kwon, Namyoung Kim, Hyung-Gi Byun, Kangwon National University, Republic of Korea; Krishna Persaud, University of Manchester, United Kingdom*

Future multimedia communication system can be developed to identify, transmit and provide odors besides voice and image. In this paper, an improved odor identification method is introduced. We present an analysis of center-gradient and a new method of using convergence parameters in training RBFN-SVD-SG (Radial Basis Function Network using Singular Value Decomposition combined with Stochastic Gradient) algorithm for odor identification. Through mathematical analysis, it was found that the steady-state weight fluctuation and large values of convergence parameter can lead to an increase of variance of center-gradient, which induces ill-behaving convergence. The proposed method of using raised-cosine functions for time-decaying convergence parameter shows faster convergence and better recognition performance.

## **TA2-P3.29 CONTRAST ENHANCEMENT IN DIRECT-PROJECTED AUGMENTED REALITY**

*Hanhoon Park, Moon-Hyun Lee, Sang-Jun Kim, Jong-Il Park, Hanyang University, Republic of Korea*

In direct-projected augmented reality, use of projector makes it possible to utilize 3-D real and large objects as displays and frees from discomforts incidental to wearing a device such as HMD. However, the resulting augmentation usually has poor depth resolution due to projectors with low contrast. In direct-projected augmented reality, the radiometric compensation, which is originally employed to recover the color properties of the projector input image in direct-projected augmented reality, seems to be another cause of decreasing the final contrast of the augmentation. In this paper, a contrast enhancement method is proposed that combines a hue- and saturation- preserving histogram equalization with the radiometric compensation in direct-projected augmented reality. The method guarantees that the radiometrically compensated color is maintained while the brightness (or intensity) contrast is enhanced. Experimental results demonstrate the validity of our method.

### **TA2-P3.30 DMB (DIGITAL MULTIMEDIA BROADCASTING) VOICE EPG APPLICATION**

*Bong-Ho Lee, So Ra Park, HeeJeong Kim, ChungHyun Ahn, Soo-In Lee, Electronics and Telecommunications Research Institute (ETRI), Republic of Korea*

Recently, mobile TV is becoming a mainstream service for the mobile broadcasting that requires lots of mobile factors such as robust transmission, high performance and easy interface and so on. As mobile broadcasting services are being highlighted, the easy interface and appropriate application platforms are increasingly being in demand. In most mobile scenarios, where users may actually share their attention with other concurrent tasks and where highly integrated devices may have very limited physical characteristics, intuitive man-machine interfaces are key factors to successful applications. The EPG, a crucial application in most digital broadcasting systems, is also not free from the easy interface and mobile factors. The conventional GUI driven EPG solutions are, sometimes, not appropriate to the mobile system, especially to the DMB targeting mobile TV and rich interactive data applications. In this paper we present voice enabled EPG application that features voice user interaction and dialog technology allowing the user to have a speech interaction with the terminal in navigating and searching any program or service. We illustrate an overall service framework addressing the content delivery and consuming architecture fitted to the DMB network. Moreover, we propose and implement an agent platform by profiling the elements of VoiceXML to make it a mobile version and extending some key elements to add EPG associated functionalities to the VoiceXML platform.

### **TA2-P4: Tuesday, July 11, 10:50 - 11:50**

#### **TA2-P4 Authentication, Data Hiding and Others (Poster)**

Time: Tuesday, July 11, 10:50 - 11:50

Place: Toronto III

Chair: Stefanos Kollias, National Technical University of Athens

#### **TA2-P4.31 REVERSIBLE IMAGE AUTHENTICATION BASED ON WATERMARKING**

*Sang-Kwang Lee, Young-Ho Suh, Electronics and Telecommunications Research Institute (ETRI), Republic of Korea; Yo-Sung Ho, Gwangju Institute of Science and Technology, Republic of Korea*

In this paper, we propose a new reversible image authentication technique based on watermarking where if the image is authentic, the distortion due to embedding can be completely removed from the watermarked image after the hidden data has been extracted. This technique utilizes histogram characteristics of the difference image and modifies pixel values slightly to embed more data than other lossless data hiding algorithm. We show that the lower bound of the PSNR (peak-signal-to-noise-ratio) values of watermarked images are 51.14 dB. Moreover, the proposed scheme is quite simple and the execution time is rather short. Experimental results demonstrate that the proposed scheme can detect any modifications of the watermarked image.

#### **TA2-P4.32 ON RESAMPLING DETECTION AND ITS APPLICATION TO DETECT IMAGE TAMPERING**

*Prasad S., K.R. Ramakrishnan, Indian Institute of Science, India*

Usually digital image forgeries are created by copy-pasting a portion of an image onto some other image. While doing so, it is often necessary to resize that pasted portion of the image to suit the sampling grid of the host image. The resampling operation changes certain characteristics of the pasted portion, which when detected serves as a clue of tampering. Here we present deterministic techniques to detect resampling, and localize that portion of the image that has been tampered with. Two of the techniques are in pixel domain and two others in frequency domain. We study the efficacy of our techniques against JPEG compression and subsequent resampling of the entire tampered image.

#### **TA2-P4.33 HYBRID TRAITOR TRACING**

*Hongxia Jin, Jeffery Lotspiech, IBM Almaden Research Center, United States*

In this paper we study the traitor tracing problem, a technique to help combat piracy of copyrighted materials. When a pirated copy of the material is observed, a traitor tracing scheme allows one to identify at least one of the real users (traitors) who participate in the construction of the pirated copy. The authors have been involved in what we believe is the first large-scale deployment of the tracing traitors approach. In this paper we shall present a key management scheme that provides the system designer flexibility, in that he/ she can choose to trace down to an individual user or only trace to a device's model. It helps effectively defend against class attack and evil manufacturer problem. It also provides a technology to protect consumers' privacy when needed. This flexibility is believed to be needed and has been adopted by content protection standard AACs for next generation DVDs.

- TA2-P4.34 A SECRET KEY BASED MULTISCALE FRAGILE WATERMARK IN THE WAVELET DOMAIN**  
*Hua Yuan, Xiao-Ping Zhang, Ryerson University, Canada*  
 The distribution of the wavelet coefficients in 2-D discrete wavelet transform (DWT) subspaces can be well described by a Gaussian mixture statistical model. In this paper, a secret key based fragile watermarking scheme is presented based on this statistical model. The Gaussian statistical model parameters are obtained by an expectation maximization (EM) algorithm and modified in a way to form special relationships for image authentication. The secret key is designed to securely embed a message bit stream, such as personal signatures or copyright logos, into a host image. Because of the secret embedding key, the new method is robust to most image tampering, even when the attackers are fully aware of the watermark embedding algorithms. Besides, the secret embedding key can be encrypted and embedded as a robust watermark into the same host image of the fragile watermarks for the benefit that the decoding of fragile watermarks only requires a single encryption key other than the image itself. The new method also has the advantage of changing only a few image data for watermark embedding and being able to distinguish some normal image operations such as compression from malicious to achieve a semi-fragile application.
- TA2-P4.35 MINIMUM DISTORTION LOOK-UP TABLE BASED DATA HIDING**  
*Xiaofeng Wang, Xiao-Ping Zhang, Ryerson University, Canada*  
 In this paper, we present a novel data hiding scheme based on the minimum distortion look-up table (LUT) embedding that achieves good distortion-robustness performance. We first analyze the distortion introduced by LUT embedding and formulate its relationship with run constraints of LUT. Subsequently, a Viterbi algorithm is presented to find the minimum distortion LUT. Theoretical analysis and numerical results show that the new LUT design achieves not only less distortion but also more robustness than the traditional LUT based data embedding schemes.
- TA2-P4.36 PROTECTION OF VIDEO LOGOS WITH RANDOMIZATION**  
*Yongdong Wu, Institute for Infocomm Research, Singapore*  
 To announce the ownership of a video programs such as TV, the owner usually embeds his logo into the programs in a visible way. Since the logo and its position are usually fixed in the video frames, an adversary can completely remove the logo from the video without video quality loss. In order to thwart this removal attack, the present paper randomly changes logo location and distorts the logo without reducing the visibility and fidelity of the logo. Our experiments show the effectiveness of the present methods.
- TA2-P4.37 CONFRONTING THE SYNCHRONIZATION PROBLEM OF SEMANTIC REGION UNDER GEOMETRIC ATTACKS**  
*Paraskevi Tzouveli, Klimis Ntalianis, Stefanos Kollias, National Technical University of Athens, Greece*  
 In this paper, an affine invariant watermarking scheme, robust to geometric attacks, is proposed and applied to face regions. Initially, face regions are unsupervisedly extracted from an initial image and a normalization procedure, invariant to geometric attacks is applied on each of these regions using a set of specific moment criteria. A spread spectrum-based DS-CDMA watermarking scheme is then used in order to provide a multi bits watermark. The multi bits watermark embedding and detection procedures are then applicable to each normalized face region. Finally, performance of the proposed face regions watermarking scheme is tested under various distortions, providing efficient and robust watermark retrieval.
- TA2-P4.38 A VIDEO SCRAMBLING SCHEME APPLICABLE TO LOCAL REGION WITHOUT DATA EXPANSION**  
*Makoto Takayama, Kiyoshi Tanaka, Shinshu University, Japan; Akio Yoneyama, Yasuyuki Nakajima, KDDI R&D Laboratories, Inc, Japan*  
 Recently, several scrambling techniques have been proposed for video digitally archived. These methods realize efficient processing by partial encryption on MPEG compressed data while keeping compatibility to MPEG format. However, they have common drawbacks expanding the entire code. Also, they have not reported on local shuffling in a frame. In this work, we propose a new video scrambling scheme on MPEG compressed domain, which shuffles a part of DC and AC coefficients by DCT in a frame. Our scheme is applicable to local region without data expansion from the original MPEG file.

## **TA2-P4.39 CONDITIONAL ACCESS TO H.264/AVC VIDEO WITH DRIFT CONTROL**

*Enrico Magli, Marco Grangetto, Gabriella Olmo, Politecnico di Torino, Italy*

In this paper we address the problem of providing conditional access to video sequences, namely, to generate a low-quality video to be used as preview, which can be decoded at full quality if a decryption key is obtained. We propose and investigate the performance of two different techniques, based on smoothing and separate encoding in the compressed domain, and motion vector perturbation. We show that these techniques are able to provide conditional access to different quality levels of H.264/AVC video with very small rate overhead, and that their combination can provide different levels of security towards malicious attacks.

## **TA2-P4.40 A PROTECTION PROCESSOR FOR MPEG-21 PLAYERS**

*Paolo Nesi, Davide Rogai, Andrea Vallotti, University of Florence, Italy*

The design and implementation of MPEG-21 players and authoring tools presents several critical points to be solved. One of the most relevant is the security level and protection processing in the players. This paper presents a solution for the realization of components in charge of enforcing Digital Rights Management in AXMEDIS tools for MPEG-21 digital content. The proposed architecture provides functionalities to create both trusted environment on the client side and dynamic protection and unprotection of digital content including digital resources and their organization and metadata. The same solution can be used to achieve the desired security level in any other MPEG-21 player or authoring tool. The architecture presented hereinafter has been adopted to enforce protection on authoring and player tools developed for the AXMEDIS IST FP6 R&D European Commission project.

## **TA2-P4.41 THE WATERMARKING FOR 3D CAD DRAWING USING LINE, ARC, 3DFACE COMPONENTS**

*Ki-Ryong Kwon, Pukyong National University, Republic of Korea; Jae-Sik Sohn, Kyungpook National University, Republic of Korea; Young Huh, Korea Electrotechnology Research Institute, Republic of Korea; Suk-Hwan Lee, TongMyong University, Republic of Korea*

Currently there has been much interested in developing the watermarking for 3D graphic data of mesh model or NURBS. However, the watermarking technique based on 3D CAD drawing leaves something to be desired. This paper proposes a watermarking technique for 3D design drawing using the components of Line, 3DFACE and ARC based on vertex that prevent the infringement of copyright from unlawfulness reproductions and distribution. By experimental result, we confirmed the invisibility of embedded watermarks as well as the robustness in geometrical attacks and file conversion to DWG, DXF, DWT and DWS.

## **TP1-L1: Tuesday, July 11, 13:10 - 14:30**

### **TP1-L1 Digital Watermarking, Data Hiding and Steganography III (Lecture)**

Time: Tuesday, July 11, 13:10 - 14:30

Place: Carmichael

Chair: Gaurav Sharma, University of Rochester

13:10 - 13:30

### **TP1-L1.1 STEGANALYSIS BASED ON MARKOV MODEL OF THRESHOLDED PREDICTION-ERROR IMAGE**

*Dekun Zou, Yun Shi, Wei Su, New Jersey Institute of Technology, United States; Guorong Xuan, Tongji University, China*

A steganalysis system based on 2-D Markov chain of thresholded prediction-error image is proposed in this paper. Image pixels are predicted with their neighboring pixels, and the prediction-error image is generated by subtracting the prediction value from the pixel value and then thresholded with a predefined threshold. The empirical transition matrixes of Markov chain along the horizontal, vertical and diagonal directions serve as features for steganalysis. Support vector machines (SVM) are utilized as classifier. The effectiveness of the proposed system has been demonstrated by extensive experimental investigation. The detection rate for Cox et al.'s non-blind spread spectrum (SS) data hiding method, Piva et al.'s blind SS method, and a generic QIM method (as embedding data rate being 0.1 bpp (bits per pixel)) are all above 90% over an image database consisting of approximately 4000 images. For generic LSB method (with various embedding data rates), our steganalysis system achieves a detection rate above 85% as the embedding data rate is 0.1 bpp and above.

13:30 - 13:50

**TP1-L1.2 ON THE DETECTION OF MULTIPLICATIVE WATERMARKS FOR SPEECH SIGNALS IN THE WAVELET AND DCT DOMAINS**

*Ramin Eslami, John Deller, Hayder Radha, Michigan State University, United States*

Blind multiplicative watermarking schemes for speech signals using wavelets and discrete cosine transform are presented. Watermarked signals are modeled using a generalized Gaussian distribution (GGD) and Cauchy probability model. Detectors are developed employing generalized likelihood ratio test (GLRT) and locally most powerful (LMP) approach. The LMP scheme is used for the Cauchy distribution, while the GLRT estimates the gain factor as an unknown parameter in the GGD model. The detectors are tested using Monte Carlo simulation and results show the superiority of the proposed LMP/Cauchy detector in some experiments.

13:50 - 14:10

**TP1-L1.3 ASYMPTOTICALLY OPTIMAL SCALAR QUANTIZERS FOR QIM WATERMARK DETECTION**

*Jean-Philippe Boyer, LSS, France; Pierre Duhamel, LSS/CNRS, France; Jacques Blanc-Talon, CTA/GIP, France*

This paper investigates asymptotically optimal scalar quantizers to address QIM watermark detection with i.i.d. host data and additive noise. False-alarm probability of detection is chosen as the cost to be minimized, keeping the embedding distortion and the miss probability upper-bounded. To avoid the intractability of false-alarm probability, Kullback distance between watermarked and non-watermarked data is adopted instead. The problem is then to seek the quantizer which maximizes the false-alarm error exponent under distortion constraint. Using Lagrange multiplier minimization, a quantizer updating Lloyd-Max-like procedure is used to solve the optimization. For experimental aspects, host data and noise have been set gaussian. In comparison with uniform or Lloyd-Max quantizers, it turns out that detection performances can be notably enhanced by using proposed application-optimized quantizers. The gain is effective even for small number  $N$  of sample at the detector input. However, this gain becomes more substantial as  $N$  grows. This also emphasises that good quantizers in terms of distortion are not suitable for detection task.

14:10 - 14:30

**TP1-L1.4 DETECTION OF LSB STEGANOGRAPHY BASED ON IMAGE SMOOTHNESS**

*Tao Zhang, Yan Zhang, Xijian Ping, Mingwu Song, Zhengzhou Information Science and Technology Institute, China*

The detection of LSB steganography is a question of common interest in the research of steganalysis techniques. In this paper, the distribution of the difference between the current pixel value and its neighborhood average pixel value is statistically modeled, and then the variance of this statistical distribution is defined as a measurement of image smoothness. Based on the analysis of the effects on the image smoothness brought by message embedding and LSB plane flipping, a new steganalytic technique capable of reliable detection of spatial LSB steganography is proposed. The algorithm can exactly estimate the amount of hidden messages and detect the existence of hidden messages embedded in the image simultaneously. Experimental results show that the proposed algorithm is effective.

**TP1-L2 Analysis and Synthesis of Speech and Audio II (Lecture)**

Time: Tuesday, July 11, 13:10 - 14:30

Place: Johnston

Chair: Sumit Basu, Microsoft Research

13:10 - 13:30

**TP1-L2.1 SPEAKER IDENTIFICATION USING A MICROPHONE ARRAY AND A JOINT HMM WITH SPEECH SPECTRUM AND ANGLE OF ARRIVAL**

*Jack Stokes, John Platt, Sumit Basu, Microsoft Research, United States*

In this paper, we present a speaker identification algorithm for a microphone array based on a first-order joint Hidden Markov Model (HMM) where the observations correspond to the angle of arrival of the speech and the speech spectrum. The goal of the research is to investigate whether including angle of arrival information improves the speaker identification error rates compared to an algorithm based on the speech spectrum only. The spectral model consists of a Gaussian Mixture Model (GMM) using Multiple Discriminant Analysis (MDA) coefficients, and the angle model includes a separate histogram for each participant. The convergence time of the joint HMM is improved by estimating the GMM for each of the meeting participants prior to the start of the meeting and initializing each participant's spectral GMM in the joint HMM to the pretrained parameter values. The performance of the algorithm is analyzed from data collected during live meetings recorded using an eight element, circular microphone array. For meetings where the participants are stationary, the results show significant improvement over a single channel speaker ID algorithms based on spectrum only.

13:30 - 13:50

**TP1-L2.2 HIERARCHICAL GENRE CLASSIFICATION FOR LARGE MUSIC COLLECTIONS**

*Stefan Brecheisen, Hans-Peter Kriegel, Peter Kunath, Alexey Pryakhin, University of Munich, Germany*

The rapid progress in digital music distribution has lead to the creation of large collections of music. There is a need for content-based music classification methods to organize these collections automatically using a given genre taxonomy. To provide a versatile description of the music content, several kinds of features like rhythm, pitch or timbre characteristics are commonly used. Taking the highly dynamic nature of music into account, each of these features should be calculated up to several hundreds of times per second. Thus, a piece of music is represented by a complex object given by several large sets of feature vectors. In this paper, we propose a novel approach for the hierarchical classification of music pieces into a genre taxonomy. Our approach is able to handle multiple characteristics of music content and achieves a high classification accuracy efficiently, as shown in our experiments performed on a real world data set.

13:50 - 14:10

**TP1-L2.3 MUSIC SIGNAL SYNTHESIS USING SINUSOID MODELS AND SLIDING-WINDOW ESPRIT**

*Anders Gunnarsson, Irene Y.H. Gu, Chalmers University of Technology, Sweden*

This paper proposes a music signal synthesis scheme that is based on sinusoid modeling and sliding-window ESPRIT. Despite widely used audio coding standards, effectively synthesizing music using sinusoid models, more suitable for harmonic rich music signals, remains an open issue. In the proposed scheme, music signals are modeled by a sum of damped sinusoids in noise. A sliding window ESPRIT algorithm is applied. A continuity constraint is then imposed for tracking the time trajectories of sinusoids in music and for removing spurious spectral peaks in order to adapt to the changing number of sinusoid contents in dynamic music. Simulations have been performed to several music signals with a range of complexities, including music recorded from banjo, flute and music with mixed instruments. The results from listening and spectrograms have strongly indicated that the proposed method is very robust for music synthesis with good quality.

14:10 - 14:30

**TP1-L2.4 PERFORMANCE ANALYSIS OF COMPRESSED-DOMAIN AUTOMATIC SPEAKER RECOGNITION AS A FUNCTION OF SPEECH CODING TECHNIQUE AND BIT RATE**

*Matteo Petracca, Antonio Servetti, Juan Carlos De Martin, Politecnico di Torino, Italy*

Compressed-domain automatic speaker recognition is based on the analysis of the compressed parameters of speech coders. The objective is to perform low-complexity on-line speaker recognition for VoIP in the compressed domain, without the need to decode or resynthesize the speech bitstream. In this paper, we present initial results in determining the recognition accuracy that can be achieved with five widely used speech coding standards. Experiments with a database of 14 speakers obtain a recognition ratio close to 100% after the analysis of 30 seconds of active speech for most of the considered speech coders and rates. In particular, the results show that performance does not strictly depend on coding rate or codec speech quality.

**TP1-L3: Tuesday, July 11, 13:10 - 14:30**

**TP1-L3 Advances in Networked Video** (Special Session)

Time: Tuesday, July 11, 13:10 - 14:30

Place: Thomson

Co-Chairs: Mary Comer, Purdue University and Xiaojun Lin, Purdue University

13:10 - 13:30

**TP1-L3.1 PEER-TO-PEER ASYNCHRONOUS VIDEO STREAMING USING SKIP LIST**

*Dan Wang, Jiangchuan Liu, Simon Fraser University, Canada*

Media distribution through application-layer overlay networks has received considerable attention recently, owing to its flexibility and readily deployable nature. On-demand streaming with asynchronous requests, and in general, with VCR-like interactions, nevertheless remains a challenging task. In this paper, we introduce the Skip List, a novel randomized and distributed structure that inherently accommodates dynamic and asynchronous clients. We demonstrate a practical skip list based streaming overlay with typical VCR operations. Our simulation results show that the skip list based overlay is highly scalable, with smooth playback for diverse interactivities, and low overheads.

13:30 - 13:50

**TP1-L3.2 DISTRIBUTED MEDIA RATE ALLOCATION IN OVERLAY NETWORKS**

*Dan Jurca, Pascal Frossard, Ecole Polytechnique Fédérale de Lausanne (EPFL), Switzerland*

We address the problem of distributed path selection and rate allocation for media streaming in overlay networks. Under the assumption that each node has only a local view of the network, we propose a distributed algorithm for joint path selection, and rate allocation, in order to minimize the end-to-end media distortion. The distributed algorithm performs iteratively, by greedy rate allocation for all incoming media flows on the outgoing links at each intermediate node. Our algorithm is shown to converge to the optimal rate allocation solution in a very small number of iterations, and to outperform heuristic distributed rate allocation mechanisms for a number of random network topologies.

13:50 - 14:10

**TP1-L3.3 VIDEO TEXTURE AND MOTION BASED MODELING OF RATE VARIABILITY-DISTORTION (VD) CURVES OF I, P, AND B FRAMES**

*Geert Van der Auwera, Martin Reisslein, Lina J. Karam, Arizona State University, United States*

We examine the bit rate variability-distortion (VD) curve of I, P, and B frames of MPEG-4 VBR encoded video sequences. We show that the concave VD curve shape at high compression ratios or large quantization scales, is influenced by both the texture and the motion information. We use linear and quadratic models for the texture and motion bits statistics and devise accurate VD curve models. The model parameters are obtained from statistics that are estimated from two encodings. This work extends our previous work on modeling the VD curve, which has applications for optimal statistical multiplexing of VBR streaming video.



#### **TP1-L3.4 AN UNBALANCED MULTIPLE DESCRIPTION CODING SCHEME FOR VIDEO TRANSMISSION OVER WIRELESS AD HOC NETWORKS**

*Li Bin, Huang Feng, Sun Lifeng, Yang Shiqiang, Tsinghua University, China*

Video transmission over wireless ad hoc networks is hampered by packet losses. Even a single packet loss may cause error propagation until an intra-coded frame is received. Indeed, packet losses greatly degrade the video quality. In this paper, we propose an Unbalanced Multiple Description Coding (UMDC) scheme over a single path which requires only one single path as additional links are difficult to be guaranteed in reality and is capable of quickly recovering from packet losses and ensuring continuous playback. The proposed scheme uses two descriptions, the High-Resolution (HR) description and the Low-Resolution (LR) one. It uses the 'peg frames' to limit error propagation in the HR description. The two descriptions can help each other recover from packet losses. The simulation results show that the proposed UMDC scheme over a single path has a comparable performance with our UMDC scheme with multiple path transmission (MPT) and has a better viewing experience than the state-of-the-art FEC-based scheme.

### **TP1-L4: Tuesday, July 11, 13:10 - 14:30**

#### **TP1-L4 P2P Multimedia III (Special Session)**

Time: Tuesday, July 11, 13:10 - 14:30

Place: Toronto I

Chair: Zhu Liu, AT&T Research Labs

13:10 - 13:30

#### **TP1-L4.1 DISCOVER: DISTRIBUTED COLLABORATIVE VIDEO RECORDER**

*Jin Li, Cheng Huang, Microsoft Research, United States*

The paper describes DISCOVER, a distributed collaborative video recorder. DISCOVER is a P2P application that combines asynchronous file sharing with synchronous on-demand media streaming. DISCOVER uses a flat entity ID space, with the entity being any of the media file, header, mega packets, index and metadata. All DISCOVER entities may be asynchronously or synchronously distributed. DISCOVER adopts a sender-driven priority based sharing protocol. If the user is on-demand viewing a media file, those packets that are to be viewed in the near future will be put on the synchronous access list, which promotes its connected peers and the peers that are indirectly connected to fulfill the distribution of the on-demand packets in high priority. By letting the peers engage in both asynchronous sharing and synchronous on-demand streaming, DISCOVER promotes the peers to remain online longer, thus improve the availability of the P2P system and the overall performance.

13:30 - 13:50

#### **TP1-L4.2 RIPPLE-STREAM: SAFEGUARDING P2P STREAMING AGAINST DOS ATTACKS**

*Wenjie Wang, University of Michigan, United States; Yongqiang Xiong, Microsoft Research Asia, China; Qian Zhang, Hong Kong University of Science and Technology, China; Sugih Jamin, University of Michigan, United States*

Compared with file-sharing and distributed hash table (DHT) network, P2P streaming is more vulnerable to denial of service (DoS) attacks because of its high bandwidth demand and stringent time requirement. This paper studies the design of DoS resilient streaming networks using credit systems. We propose a novel framework--ripple-stream--to improve DoS resilience of P2P streaming. Ripple-stream leverages existing credit systems to introduce credit constraints in overlay construction such that malicious nodes are pushed to the fringe of overlays. Combining credit constraints with overlay optimization techniques, ripple-stream can achieve both DoS resilience and overlay efficiency.

13:50 - 14:10

**TP1-L4.3 DESIGN AND VERIFICATION OF COMMUNICATION PROTOCOLS FOR PEER-TO-PEER MULTIMEDIA SYSTEMS**

*Senem Velipasalar, Chang-Hong Lin, Jason Schlessman, Wayne Wolf, Princeton University, United States*

This paper addresses issues pertaining to the necessity of utilizing formal verification methods in the design of protocols for peer-to-peer multimedia systems. These systems require sophisticated communication protocols, and these protocols require verification. We discuss two sample protocols designed for two distinct peer-to-peer computer vision applications, namely multi-object multi-camera tracking and distributed gesture recognition. We present simulation and verification results for these protocols, obtained by using the SPIN verification tool, and discuss the importance of verifying the protocols used in peer-to-peer multimedia systems.

14:10 - 14:30

**TP1-L4.4 PERSONAL MEDIA ALERT SYSTEMS: PERSONALIZATION AND DISSEMINATION OF BROADCAST CONTENT WITH A P2P MICROPAYMENT SCHEME**

*Yih-Farn Chen, David Gibbon, Zhu Liu, Behzad Shahraray, Bin Wei, AT&T Labs - Research, United States*

Media consumers who are overwhelmed by multitudinous news content demand a system that would sift through tens or hundreds of broadcast TV channels on a daily basis to capture the most important clips that match users' interests and deliver these personalized clips for easy viewing on a typical TV or home PC. In this paper, we present a personal media alert system that extracts video segments for the interested users. In addition, alerts with pointers to content stored on the home server can be sent to mobile users who are authorized to access the content at home. To make content sharing feasible and scalable, we also propose a P2P payment scheme that would allow a consumer to have access to media clips aggregated on other consumers' personal media systems.

**TP1-L5: Tuesday, July 11, 13:10 - 14:30**

**TP1-L5 Joint Source and Channel Coding for Wireless Multimedia (Lecture)**

Time: Tuesday, July 11, 13:10 - 14:30

Place: Toronto II

Chair: Anthony Vetro, Mitsubishi Electric Research Lab

13:10 - 13:30

**TP1-L5.1 JOINT SOURCE-CHANNEL DECODING OF MULTIPLE DESCRIPTION QUANTIZED AND VARIABLE LENGTH CODED MARKOV SEQUENCES**

*Xiaohan Wang, Xiaolin Wu, McMaster University, Canada*

This paper proposes a framework for joint source-channel decoding of Markov sequences that are encoded by an entropy coded multiple description quantizer (MDQ), and transmitted via a lossy network. This framework is particularly suited for lossy networks of inexpensive energy-deprived mobile source encoders. Our approach is one of maximum a posteriori probability (MAP) sequence estimation that exploits both the source memory and the correlation between different MDQ descriptions. The MAP problem is modeled and solved as one of the longest path in a weighted directed acyclic graph. For MDQ-compressed Markov sequences impaired by both bit errors and erasure errors, the proposed joint source-channel MAP decoder can achieve 5dB higher SNR than the conventional hard-decision decoder. Furthermore, the new MDQ decoding technique unifies the treatments of different subsets of the K descriptions available at the decoder, circumventing the thorny issue of requiring up to  $2^K-1$  MDQ side decoders.

13:30 - 13:50

**TP1-L5.2 A NEW JOINT SOURCE-CHANNEL MODELING APPROACH FOR ADAPTIVE FEC CODE RATE DECISION**

*Yo-Won Jeong, Korea Advanced Institute of Science and Technology, Republic of Korea; Jae Cheol Kwon, KT, Republic of Korea; Jae-kyoon Kim, Kyu-Ho Park, Korea Advanced Institute of Science and Technology, Republic of Korea*

We propose a joint source-channel modeling approach for adaptively determining the quality-optimal FEC code rate. Our objective is to obtain the optimal video quality in the receiver, while taking time-varying packet loss into consideration. First, we define the loss threshold set as the set of packet loss probabilities in which the code rate have to be adjusted to maintain the maximum video quality. Proposed model can estimate the loss threshold set efficiently. Then, whenever there is a change of the packet loss status in the channel, video sender can always find the optimal code rate on-the-fly by using the estimated loss threshold set. Simulation results show that the proposed method can determine the near-optimal code rate in joint source-channel coding.

13:50 - 14:10

**TP1-L5.3 DYNAMIC RATE CONTROL FOR JPEG 2000 TRANSCODING**

*Derek Schwenke, Anthony Vetro, Mitsubishi Electric Research Laboratories, United States; Toshihiko Hata, Naoki Kuwahara, Mitsubishi Electric Corporation, Japan*

This paper describes a rate control algorithm for a streaming video system that dynamically transcodes stored JPEG 2000 frames. The proposed algorithm is designed to improve overall quality over a static rate control method by increasing bandwidth utilization, while satisfying buffer constraints and maintaining consistent quality over time. Simulation results confirm the effectiveness of the proposed algorithm in terms of both objective measures and subjective evaluation.

14:10 - 14:30

**TP1-L5.4 CONTROLLED COMPLEXITY MAP DECODING OF CABAC ENCODED DATA**

*Salma Ben-Jamaa, Michel Kieffer, Pierre Duhamel, LSS, France*

In this paper, we present a joint source-channel decoding technique based on exact MAP estimation for data encoded by CABAC (Context-based Adaptive Binary Arithmetic Coding) in standards like H.264/AVC. Soft decoding is put at work using an improved sequential decoding technique, achieving a trade-off between complexity and efficiency of the proposed algorithms. Error detection is realized by exploiting the binarization scheme and redundancy left in the code string, so that CABAC compression efficiency is preserved and no additional redundancy is compulsory.

**TP1-L6: Tuesday, July 11, 13:10 - 14:30**

**TP1-L6 Advanced Media Coding I (Lecture)**

Time: Tuesday, July 11, 13:10 - 14:30

Place: Varley

Chair: Gene Cheung, Hewlett-Packard Laboratories

13:10 - 13:30

**TP1-L6.1 AN OPTIMIZED MAPPING ALGORITHM FOR CLASSIFIED VIDEO TRANSMISSION WITH THE H.264 FLEXIBLE MACROBLOCK ORDERING**

*Sio Kei Im, A. J. Pearmain, Queen Mary University of London, United Kingdom*

A new mapping algorithm is proposed for categorization of frames' macroblocks into two (or more) classes in flexible macroblock ordering to give error resilient video transmission. The successful transmission of macroblock data not only enhances the quality of the associated pixels, but also improves the quality of the adjacent lost macroblocks by improving the efficiency of error concealment. Therefore, in our scheme, by carefully modeling the decoder error concealment algorithm at the encoder side, we classify the macroblocks according to their eventual influence on picture quality. Within a limited bit rate budget, we employ an optimization algorithm to select the best group of high-priority macroblocks. We show that prioritized transmission of the more important macroblock group will improve the video quality in error situations where our mapping algorithm outperforms the default mappings of the H.264 codec.

13:30 - 13:50

**TP1-L6.2 COMPLEXITY SCALABLE 2:1 RESOLUTION DOWNSCALING MPEG-2 TO WMV  
TRANSCODER WITH ADAPTIVE ERROR COMPENSATION**

*Guobin Shen, Microsoft Research Asia, China; Yuwen He, Panasonic Laboratories, Singapore; Wanyong Cao, Shipeng Li, Microsoft Research Asia, China*

In this paper, we focus on 2 : 1 spatial resolution downscaling transcoding from MPEG-2 to WMV. We propose two architectures (for sequences with or without B-frames respectively) that are unique in their complexity scalability and efficient control over the drifting error, which in return provide a flexible mechanism to achieve desired tradeoff between the complexity and the quality. We achieve resolution downscaling completely in the DCT domain and show that the standard IDCT (as in all the MPEG series standards) can be merged with other DCT-like transform (e.g., the integer transform in WMV) with proper one-time per-element scaling. Extensive experimental results verified the effectiveness of proposed structures against several design objectives such as complexity scalability and performance tradeoffs.

13:50 - 14:10

**TP1-L6.3 COMPLEXITY-DISTORTION OPTIMIZED MOTION ESTIMATION ALGORITHM WITH FINE-GRANULAR SCALABLE COMPLEXITY**

*Li Zhang, Wen Gao, Chinese Academy of Sciences, China*

Video encoding now is being implemented in various computing platforms with different computing capability, the requirement on the encoding complexity is also different according to different applications. As the most computation-intensive part of video encoding, the ME (motion estimation) should have a scalable complexity. This paper proposes a ME algorithm with fine-granular scalable complexity, a more important feature of the proposed algorithm is that it seeks for the complexity-distortion optimization. The given computation budget will be allocated to each MB (macroblock) in one frame. Each MB will consume its allocated computation by a hybrid search pattern. Experimental results show that the proposed algorithm can get a better computation-distortion performance than the existing ME algorithms.

14:10 - 14:30

**TP1-L6.4 TWO-LAYER IMAGE RESIZING FOR SCALABLE CODEC**

*Wang Ci, Xue Ping, Nanyang Technological University, Singapore; Weisi Lin, Institute for Infocomm Research, Singapore*

Resizing compressed images is often needed for different display, storage and transmission. From the analysis of energy reserving capacity for different basis vectors, we propose a two-layer downsizing scheme, in which the elementary layer carries the low frequency components of the image, while some high frequency details can be reasonably recovered from the enhancement layer. The new method provides a solution for low definition images scalable transmission, at the same time, gives its potential to rationally recover the high definition image. Experimental results show that the new scheme exhibits better picture quality than some current downsizing approaches, but its computational cost is comparable with the others. The conception of this subband decomposition can be further used to other ratio image resizing.

**TP1-P1: Tuesday, July 11, 13:30 - 14:30**

**TP1-P1 Event and Activity Recognition II (Poster)**

Time: Tuesday, July 11, 13:30 - 14:30

Place: Toronto III

Chair: Mark Drew, Simon Fraser University

**TP1-P1.1 A BEHAVIOURAL APPROACH TO PERSON RECOGNITION**

*Federico Matta, Jean-Luc Dugelay, Eurecom Institute, France*

This paper describes a new approach for identity recognition using video sequences. While most image and video recognition systems discriminate identities using physical information only, our approach exploits the behavioural information of head dynamics; in particular the displacement signals of few head features directly extracted in the image plane. Due to the lack of standard video database, identification and verification scores have been obtained using a small collection of video sequences: the results for this new approach are nevertheless promising.

**TP1-P1.2 USING DECISION-TREE TO AUTOMATICALLY CONSTRUCT LEARNED-HEURISTICS FOR EVENTS CLASSIFICATION IN SPORTS VIDEO**

*Dian Tjondronegoro, Queensland University of Technology, Australia; Yi-Ping Phoebe Chen, Deakin University, Australia*

Automatic events classification is an essential requirement for constructing an effective sports video summary. It has become a well-known theory that the high-level semantics in sport video can be “computationally interpreted” based on the occurrences of specific audio and visual features which can be extracted automatically. State-of-the-art solutions for features-based event classification have only relied on either manual-knowledge based heuristics or machine learning. To bridge the gaps, we have successfully combined the two approaches by using learning-based heuristics. The heuristics are constructed automatically using decision tree while manual supervision is only required to check the features and highlight contained in each training segment. Thus, fully automated construction of classification system for sports video events has been achieved. A comprehensive experiment on 10 hours video dataset, with five full-match soccer and five full-match basketball videos, has demonstrated the effectiveness/robustness of our algorithms.

**TP1-P1.3 THE SEMANTIC PATHFINDER FOR GENERIC NEWS VIDEO INDEXING**

*Cees Snoek, Marcel Worring, Jan-Mark Geusebroek, Dennis Koelma, Frank Seinstra, Arnold Smeulders, University of Amsterdam, Netherlands*

This paper presents the semantic pathfinder architecture for generic indexing of video archives. The pathfinder automatically extracts semantic concepts from video based on the exploration of different paths through three consecutive analysis steps, closely linked to the video production process, namely: content analysis, style analysis, and context analysis. The virtue of the semantic pathfinder is its learned ability to find a best path of analysis steps on a per-concept basis. To show the generality of this indexing approach we develop detectors for a lexicon of 32 concepts and we evaluate the semantic pathfinder against the 2004 NIST TRECVID video retrieval benchmark, using a news archive of 64 hours. Top ranking performance indicates the merit of the semantic pathfinder.

**TP1-P1.4 EXPLOITING TOPIC THREAD STRUCTURES IN A NEWS VIDEO ARCHIVE FOR THE SEMI-AUTOMATIC GENERATION OF VIDEO SUMMARIES**

*Ichiro Ide, Nagoya University, Japan; Hiroshi Mo, Norio Katayama, Shin'ichi Satoh, National Institute of Informatics, Japan*

We propose a method that semi-automatically composes video stories by connecting individual stories in a news video archive along a topic-based semantic structure, namely the topic thread. We introduce the methods to realize the composition, namely, story segmentation, topic threading and clustering. We then evaluate the proposed approach based on preliminary tests. Since the thread structure reflects the development of topics in the real-world, we believe that the composed news video story should be effective for the user to gain a deeper understanding of the current topic of interest.

**TP1-P1.5 REGION ENHANCED SCALE-INVARIANT SALIENCY DETECTION**

*Feng Liu, Michael Gleicher, University of Wisconsin - Madison, United States*

Saliency measures the low-level stimuli to human vision, and serves as an alternative to semantic image understanding. This paper presents a region enhanced scale-invariant saliency detection method. Our method constructs a scale-invariant saliency map from an image, segments the image into regions, and enhances the saliency map with the region information. Compared with previous methods, our method has advantages in providing robust scale-invariant saliency, giving meaningful region information for applications, and eliminating misleading high-contrast edges.

**TP1-P1.6 RECOGNIZING COMMERCIALS IN REAL-TIME USING THREE VISUAL DESCRIPTORS AND A DECISION-TREE**

*Ronald Glasberg, Cengiz Tas, Thomas Sikora, Technical University Berlin, Germany*

We present a new approach for classifying mpeg-2 video sequences as ‘commercial’ or ‘non-commercial’ by analyzing specific color, texture and motion features of consecutive frames in real-time. This is part of the well-known video-genre-classification problem, where popular TV-broadcast genres like cartoon, commercial, music, news and sports are studied. Such applications have also been discussed in the context of MPEG-7. In our method the extracted features from three visual descriptors are logically combined using a decision tree to produce a reliable recognition. The results demonstrate a high identification rate based on a large collection of 200 representative video sequences (40 ‘commercials’ and 4\*40 ‘non-commercials’) gathered from free digital TV-broadcasting in Germany.

#### **TP1-P1.7 ENHANCED SEMI-SUPERVISED LEARNING FOR AUTOMATIC VIDEO ANNOTATION**

*Meng Wang, University of Science and Technology of China, China; Xian-Sheng Hua, Microsoft Research Asia, China; Li-Rong Dai, Yan Song, University of Science and Technology of China, China*

For automatic semantic annotation of large-scale video database, the insufficiency of labeled training samples is a major obstacle. General semi-supervised learning algorithms can help solve the problem but the improvement is limited. In this paper, two semi-supervised learning algorithms, self-training and co-training, are enhanced by exploring the temporal consistency of semantic concepts in video sequences. In the enhanced algorithms, instead of individual shots, time-constraint shot clusters are taken as the basic sample units, in which most mis-classifications can be corrected before they are applied for re-training, thus more accurate statistical models can be obtained. Experiments show that enhanced self-training/co-training significantly improves the performance of video annotation.

#### **TP1-P1.8 CLASSIFIER OPTIMIZATION FOR MULTIMEDIA SEMANTIC CONCEPT DETECTION**

*Sheng Gao, Qibin Sun, Institute for Infocomm Research, Singapore*

In this paper, we present an AUC (i.e., the Area Under the Curve of Receiver Operating Characteristics (ROC)) maximization based learning algorithm to design the classifier for maximizing the ranking performance. The proposed approach trains the classifier by directly maximizing an objective function approximating the empirical AUC metric. Then the gradient descent based method is applied to estimate the parameter set of the classifier. Two specific classifiers, i.e. LDF (linear discriminant function) and GMM (Gaussian mixture model), and their corresponding learning algorithms are detailed. We evaluate the proposed algorithms on the development set of TRECVID'05 for semantic concept detection task. We compare the ranking performances with other classifiers trained using the ML (maximum likelihood) or other error minimization methods such as SVM. The results of our proposed algorithm outperform ML and SVM on all concepts in terms of its significant improvements on the AUC or AP (average precision) values. We therefore argue that for semantic concept detection, where ranking performance is much interested than the classification error, the AUC maximization based classifiers are preferred.

#### **TP1-P1.9 SEMANTIC LABELING OF MULTIMEDIA CONTENT CLUSTERS**

*Jelena Tešić, John R. Smith, IBM T. J. Watson Research Center, United States*

In this paper we present a novel approach for labeling clusters of multimedia content that leverages supervised classification techniques in conjunction with unsupervised clustering. Recent research has produced significant results for automatic tagging of video content such as broadcast news. For example, powerful techniques have been demonstrated in the context of the NIST TRECVID video retrieval benchmark. However, the information needs of users typically span a range of semantic concepts. One of the challenges of these multimedia retrieval systems is to organize the video data in such a way that allows the user to most efficiently navigate the semantic space for the video data set. One important tool for video data organization is clustering. However, clustering results cannot be leveraged effectively when they are not labeled. We propose to build on clustering by aggregating the automatically tagged semantics. We propose and compare four techniques for labeling the clusters and evaluate the performance compared to human labeled ground-truth. We present examples of the cluster labeling results obtained on the BBC stock shots from the TRECVID-2005 video data set.

#### **TP1-P1.10 DETECTING HUMAN ACTION IN ACTIVE VIDEO**

*Hao Jiang, Ze-Nian Li, Mark S. Drew, Simon Fraser University, Canada*

We propose a novel scheme to detect human actions in active video. Active videos such as movies or sports broadcasting are taken purposively by “clever” photographers. They are object and action oriented and usually involve complex camera motions. Detecting actions in active videos is both important and challenging. We study a three-step scheme to detect complex human actions in such videos. The proposed method first locates potential objects and removes clutter with a composite filter scheme. The detected object candidates in successive frames are then associated to form object trajectories based on a consistent labeling formulation, and solved with belief propagation. Finally, specific human actions are detected in video with a linear programming matching approach that can efficiently deal with matching problems having a large target point set. The proposed method has been successfully applied in action detection for general videos and TV hockey games.

**TP1-P2 Multimedia Networking (Poster)**

Time: Tuesday, July 11, 13:30 - 14:30

Place: Toronto III

Chair: Tinh Nguyen, Oregon State University

**TP1-P2.11 STREAMING OF SCALABLE VIDEO FROM MULTIPLE SERVERS USING RATELESS CODES**  
*Jean-Paul Wagner, Jacob Chakareski, Pascal Frossard, Ecole Polytechnique Fédérale de Lausanne (EPFL), Switzerland*

This paper presents a framework for efficiently streaming scalable video from multiple servers over heterogeneous network paths. We propose to use rateless codes, or Fountain codes, such that each server acts as an independent source, without the need to coordinate its sending strategy with other servers. In this case, the problem of maximizing the received video quality and minimizing the bandwidth usage, is simply reduced to a rate allocation problem. We provide an optimal solution for an ideal scenario where the loss probability on each server-client path is exactly known. We then present a heuristic-based algorithm, which implements an unequal error protection scheme for the more realistic case of imperfect knowledge of the loss probabilities. Simulation results finally demonstrate the efficiency of the proposed algorithm, in distributed streaming scenarios over lossy channels.

**TP1-P2.12 SUBSTREAM ALLOCATION IN LAYERED P2P STREAMING**  
*Yifeng He, Ivan Lee, Ling Guan, Ryerson University, Canada*

In layered P2P streaming system, how to allocate the number of the copies for each layer is a challenging problem. Different allocation has different impact on the system performance. In this paper, we present a substream allocation scheme in layered P2P streaming. The proposed allocation scheme is adaptive to the request rate and number of the qualified peers. The simulation results show that the proposed allocation scheme enables the system to achieve an overall better quality compared to the general allocation schemes with fixed allocation percentages. In addition, the proposed allocation scheme can accelerate the growth of peer population in the initial stage of hybrid P2P streaming systems.

**TP1-P2.13 FORMAT-INDEPENDENT MULTIMEDIA STREAMING**  
*Joseph Thomas-Kerr, Ian Burnett, Christian Ritz, University of Wollongong, Australia*

The Bitstream Binding Language (BBL) is a new technology developed by the authors and being standardized by MPEG, which describes how multimedia content and metadata can be mapped onto streaming formats. This paper describes a particular application of BBL – format-independent multimedia streaming. This means that streaming servers no longer require additional software modules in order to support new content formats as they are introduced. Instead, the server requires only a BBL description of the mapping between the content format and the stream, and any content in the new format may then be delivered by the streaming server. This approach is validated using the H.264/AVC format as an example, and performance data are provided.

**TP1-P2.14 ON-DEMAND PARTIAL SCHEMA DELIVERY FOR MULTIMEDIA METADATA**  
*Stephen Davis, Ian Burnett, University of Wollongong, Australia*

XML is a popular approach to interoperable exchange of Multimedia metadata between a wide range of devices. This paper explores extending the use of the Remote XML Exchange Protocol (previously proposed by the authors) as a mechanism to provide efficient interaction with complex Multimedia XML documents and their associated schemas. This is particularly applicable to users with limited application complexity devices and/or limited bandwidth connections. Many XML documents do not fully utilize all the information present in a given schema; thus, users download substantial redundant information for the current application. This paper introduces the use of RXEP for the transmission of small, relevant schema sections. The paper investigates the advantages of schema retrieval using RXEP in terms of the bandwidth saved.



## **TP1-P2.15 USING RATE-DISTORTION METRICS FOR REAL-TIME INTERNET VIDEO STREAMING WITH TCP**

*Antonios Argyriou, Georgia Institute of Technology, United States*

In this paper we explore use of a new rate-distortion metric for optimizing real-time Internet video streaming with the transmission control protocol (TCP). The basic idea is to combat packet delays caused by TCP retransmissions that are essentially interpreted as errors by the streaming application. To this aim, we develop an analytical model of the expected video distortion at the decoder with respect to the expected latency for TCP packets, the channel state, and the error concealment method at the receiver. This metric is exploited with the design of a new algorithm for rate-distortion optimized encoding mode selection for video streaming with TCP (RDOMS-TCP). Real-time video streaming experiments show considerable improvement in PSNR in the range of 2 db over currently proposed TCP-based streaming mechanisms.

## **TP1-P2.16 RTP AND THE DATAGRAM CONGESTION CONTROL PROTOCOL**

*Colin Perkins, University of Glasgow, United Kingdom; Ladan Gharai, University of Southern California, United States*

We describe how the new Datagram Congestion Control Protocol (DCCP) can be used as a bearer for the Real-time Transport Protocol (RTP) to provide a congestion controlled basis for networked multimedia applications. This is a step towards deployment of congestion control for such applications, necessary to ensure the future stability of the best-effort network if high-bandwidth streaming and IPTV services are to be deployed outside of closed QoS-managed networks.

## **TP1-P2.17 VIDEO TRANSPORT OVER MULTI HOP DIRECTIONAL WIRELESS NETWORKS**

*Hong Man, Yang Li, Stevens Institute of Technology, United States; Xinhua Zhuang, University of Missouri, United States*

Exploiting directional antenna technology in wireless networks has become an attractive option because of the potential capacity increase through spatial reuse. Such development apparently has profound impact to wireless multimedia applications. While the advantages of directional transmission in single hop networks (e.g. WLANs and cellular systems) have been evident, it has also been recognized that multi hop directional networks may suffer long transmission delays and frequent link breakages if conventional routing protocols (e.g. AODV and DSR) are used. This is mainly due to the unique feature of directional transmission commonly known as "node deafness". To address this problem, we propose a multi-path routing scheme which focuses more on minimizing per-hop delays instead of traditional route length. We further propose a differentiated service (DiffServ) framework for layered video transport over multi hop directional network using QoS aware multi-path routing as well as class based queuing (CBQ). Both network layer and application layer simulation results are presented to demonstrate the effectiveness of the proposed video transport scheme over directional networks.

## **TP1-P2.18 DISTRIBUTED STREAMING VIA PACKET PARTITIONING**

*Jacob Chakareski, Pascal Frossard, Ecole Polytechnique Fédérale de Lausanne (EPFL), Switzerland*

We propose a system for adaptive streaming from multiple servers to a single receiver over separate network paths. Based on incoming packets, the receiver estimates the available bandwidth on every path and returns this information to the servers. An optimization algorithm is designed that enables the servers to independently partition the media packets among them according to the bandwidth information and such that the resulting video quality at the receiver is maximized. To this end, the algorithm takes advantage of a source pruning technique that preprocesses the media stream ahead of time. Simulation results demonstrate that the proposed streaming framework provides superior performance over a conventional transmission scheme that performs proportional packet scheduling based only on the available network bandwidth. Due to its low-complexity aspect, the framework is suitable for practical implementations of adaptive and efficient distributed streaming systems.

**TP1-P2.19 A NOVEL DISTRIBUTED AND PRACTICAL INCENTIVE MECHANISM FOR PEER TO PEER LIVE VIDEO STREAMING**

*Yun Tang, Lifeng Sun, Meng Zhang, Shiqiang Yang, Yuzhuo Zhong, Tsinghua University, China*

The successful deployment of peer-to-peer (P2P) live video streaming systems has practically demonstrated that it can scale to reliably support a large population of peers. However, peers, representing rational end users, tend to be non-cooperative when it comes to the duty rather than the self-interests, running counter to the fundamental design philosophy of P2P concept. The objective of this paper is to investigate the problem of encouraging users to balance what they take from the system with what they contribute. We first make a statistical analysis to the service logs of a practical P2P live video streaming system and reveal intrinsic characteristic of users' online duration. Second, we thus propose a novel incentive mechanism based on the composite contributions which consist of two objective metrics, i.e. the online duration and effective upstream traffic. This mechanism offers service differentiation to users with different contributions and has some desirable properties: (1) distributed nature upon gossip-based overlay structure and (2) practical oriented evaluation criteria. The experiment results over PlanetLab further verify its effectiveness.

**TP1-P2.20 DETECTING MALICIOUS HOSTS IN THE PRESENCE OF LYING HOSTS IN PEER-TO-PEER STREAMING**

*Xing Jin, S.-H. Gary Chan, W.-P. Ken Yiu, Hong Kong University of Science and Technology, Hong Kong SAR of China; Yongqiang Xiong, Microsoft Research Asia, China; Qian Zhang, Hong Kong University of Science and Technology, Hong Kong SAR of China*

Current peer-to-peer (P2P) streaming systems often assume that hosts are cooperative. However, this may not be true in the open environment of the Internet. In this paper, we discuss how to detect malicious hosts (e.g., with attacking actions and abnormal behavior) based on their history performance. In our system, each host monitors the performance of its neighbor(s) and reports this to a server. Based on the reports, the server computes host reputation with hosts of low reputation being malicious. A problem is that hosts may lie by submitting forged reports to the server. We hence formulate the reputation computing problem in the presence of lying hosts as a minimization problem and solve it by the traditional Levenberg-Marquardt algorithm. Simulation results show that our scheme can efficiently detect malicious hosts with high accuracy.

**TP1-P3: Tuesday, July 11, 13:30 - 14:30**

**TP1-P3 Face Detection (Poster)**

Time: Tuesday, July 11, 13:30 - 14:30

Place: Toronto III

Chair: Jie Wang, University of Toronto

**TP1-P3.21 DETECTION AND RESTORATION OF OCCLUSIONS FOR 3D FACE RECOGNITION**

*Alessandro Colombo, Claudio Cusano, Raimondo Schettini, University of Milano, Bicocca, Italy*

This paper presents an innovative restoration strategy which allows for an effective recognition of 3D faces, even when they are partially occluded by unforeseen, extraneous objects such as scarves, hats, glasses, and so on. First, the occluded regions are detected by considering their effects on the projections of the faces in a suitable face space; the non-occluded regions are then used to restore the missing information. Any recognition algorithm can be applied as usual to restored faces. This restoration strategy led to very satisfactory results on a test set of 52 three-dimensional faces presenting various kinds of occlusions.

**TP1-P3.22 A SIMPLE AND ACCURATE COLOR FACE DETECTION ALGORITHM IN COMPLEX BACKGROUND**

*Yu-Ting Pai, Shang-Jang Ruan, Mon-Chau Shie, Yi-Chi Liu, National Taiwan University of Science and Technology, Taiwan*

Human face detection plays an important role in many applications such as video surveillance, face recognition, and face image database management. This paper describes a fast face detection algorithm with accurate results. We use lighting compensation to improve the performance of color-based scheme, and reduce the computation complexity of feature-based scheme. Our method is effective on facial variations such as dark/bright vision, close eyes, open mouth, a half-profile face, and pseudo faces. It is worth stressing that our algorithm can also discriminate cartoon and human face correctly. The experimental results show that our approach can detect a frame in 111 msec with the 92.3% detection rate.

**TP1-P3.23 FACE RECOGNITION USING ENERGY PROBABILITY IN DCT DOMAIN**

*Jean Choi, University of Science and Technology, Republic of Korea; Yun-Su Chung, Ki-Hyun Kim, Jang-Hee Yoo, Electronics and Telecommunications Research Institute (ETRI), Republic of Korea*

In this paper, we propose a novel feature extraction method for face recognition. This method is based on Discrete Cosine Transform (DCT), Energy Probability (EP), and Linear Discriminant Analysis (LDA). We define an energy probability as magnitude of effective information. It is used to create a frequency mask in DCT domain. Our method consists of three steps. First, the spatial domain of face images is transformed into the frequency domain called DCT domain. Second, for energy probability is applied on DCT domain which acquires from face image, dimension reduction of data and optimization of valid information. At last, in order to obtain the most significant feature of face images, LDA is applied to the extracted data using frequency mask. Our experimental results show that the proposed method improves on the dimension reduction of feature space and the face recognition over the previously proposed methods.

**TP1-P3.24 AN IMPROVED APPROACH TO THE LINE-BASED FACE RECOGNITION**

*Anlong Ming, Huadong Ma, Beijing University of Posts and Telecommunications, China*

The line-based face recognition method is distinguished by its features, but its development and application is limited to some inherent drawbacks. This paper propose a method for decreasing the influence under variable illumination intensity by using the line-based singular value (LSV) feature vector instead of image gray-level value to calculate "distance" between two lines. We prove that our method is invariant to the illumination intensity. Finally, we suggest a distributed computing algorithm using grid computing to solve the multiscale computation. Experimental results show our approach is effective.

**TP1-P4: Tuesday, July 11, 13:30 - 14:30**

**TP1-P4 Digital Watermarking (Poster)**

Time: Tuesday, July 11, 13:30 - 14:30

Place: Toronto III

Chair: Borko Furht, Florida Atlantic University

**TP1-P4.25 ROBUST WATERMARKING IN THE WIGNER DOMAIN**

*Mahmood Al-khassaweneh, Selin Aviyente, Michigan State University, United States*

In this paper, a new watermarking scheme in the joint time-frequency domain is introduced. Wigner distribution is used to transform an image into the spatial-spectral domain. The proposed method selects the time-frequency cells to be watermarked based on the particular image's energy distribution in the joint domain. This approach ensures the imperceptibility of the embedded watermark. It is shown that embedding in the time-frequency domain is equivalent to a nonlinear embedding function in the spatial domain. A corresponding watermark detection algorithm is also introduced. The performance of the proposed watermarking algorithm under possible attacks, such as noise, re-sampling, rotation, filtering, and JPEG compression is illustrated.

#### **TP1-P4.26 REAL-TIME VIDEO WATERMARKING BASED ON EXTENDED M-SEQUENCES**

*Fuhao Zou, Zhengding Lu, Hefei Ling, Yanwei Yu, Huazhong University of Science and Technology, China*

This paper presents an oblivious real-time video watermarking scheme in which extended m-sequences is selected as watermark pattern. The good balance property of extended m-sequences is exploited to generate a stable reference point, which can be used to accurately compute modified amount between the watermarked and the original video in the case of original's absence, and furthermore, it helps to improve the robustness of watermarking scheme. To satisfy the real-time requirement, the processes of watermarking embedding and detection are directly operated in the variable length codeword (VLC) domain. To acquire better visual quality, the human visual system (HVS) is used to control the modification strength. Under the premise of assuring better visual quality and real-time requirements, our scheme is more robust than the schemes proposed by Lu and Frank et al..

#### **TP1-P4.27 A PIXEL-BASED ROBUST IMAGE WATERMARKING SYSTEM**

*Wenming Lu, Wanqing Li, Rei Safavi-Naini, Philip Ogunbona, University of Wollongong, Australia*

Robust image watermarking systems are required to be resistant to geometric attacks in addition to common image processing tasks, such as JPEG compression. However, robustness against geometric attacks, including rotation, scaling and translation, still remains one of the most challenging research topics in image watermarking. We propose a new pixel-based watermarking system in which a binary logo is embedded, a bit per pixel, in the pixel domain of an image. The encoder of the proposed system is based on a sliding window embedding scheme that applies the pseudo local average quantization index modulation (QIM), to achieve geometric attack robustness. The decoder employs a maximum a posteriori (MAP) estimation formulation supported by Markov random field model of the watermark to achieve robust decoding. Additionally, we demonstrate that the proposed scheme is also robust against possible watermark removal due to JPEG compression.

#### **TP1-P4.28 DISCRETE POLYNOMIAL TRANSFORM FOR DIGITAL IMAGE WATERMARKING APPLICATION**

*Lam Le, Sridhar Krishnan, Behnaz Ghoraani, Ryerson University, Canada*

In this study, we propose a new way to detect the image watermark messages modulated as linear chirp signals using discrete polynomial phase Phase Transform (DPT). The spread spectrum image watermarking algorithm embeds linear chirps as watermark messages. The phase of the chirp represents watermark message such that each phase corresponds to a different message. We extract the watermark message using a phase detection algorithm based on Discrete Polynomial Phase Transform (DPT). The DPT models the signal as polynomial and uses ambiguity function to estimate the signal parameters. The proposed method not only detects the presence of watermark, but also extracts the embedded watermark bits and ensures the message is received correctly. The robustness of the proposed detection scheme has been evaluated using checkmark benchmark attacks, and we found a guaranteed maximum bit error rate of 15%, with which watermark message is correctly detected using DPT.

#### **TP1-P4.29 A NONNEGATIVE MATRIX FACTORIZATION SCHEME FOR DIGITAL IMAGE WATERMARKING**

*Mohammadreza Ghaderpanah, Abdessamad Ben Hamza, Concordia University, Canada*

We present a new scheme for digital watermarking and secure copyright protection of digital images using nonnegative matrix factorization and singular value decomposition approaches. The proposed method improves the performance of the data embedding system effectively, and it is resistant to a variety of intentional attacks and normal visual processes. The experimental results clearly illustrate the much improved performance of the proposed watermarking methodology in comparison with existing techniques.

#### **TP1-P4.30 VIDEO WATERMARKING BASED ON NEURAL NETWORKS**

*Maher El'arbi, Chokri Ben Amar, REGIM, Tunisia; Henri Nicolas, LABRI, France*

In this paper, we propose a novel digital video watermarking scheme based on neural network and multi resolution motion estimation. Embedding and extraction of the watermark are based on the relationship between a wavelet coefficient and its neighbour's. A neural network is given to memorize the relationships between coefficients in a 3x3 block of the image. In addition, a multi resolution motion estimation algorithm is adopted to preferentially allocate the watermark to coefficients containing motion. Experimental results show that embedding watermark where picture content is moving is less perceptible. Further, it shows that the proposed scheme is robust against common video processing

**TP1-P4.31 A ROBUST IMAGE WATERMARKING SCHEME BASED ON THE ALPHA-BETA SPACE**

*Pedro Martins, Paulo Carvalho, University of Coimbra, Portugal*

A robust image watermarking scheme relying on an affine invariant embedding domain is presented. The invariant space is obtained by triangulating the image using affine invariant interest points as vertices and performing an invariant triangle representation with respect to affine transformations based on the barycentric coordinates system. The watermark is encoded via quantization index modulation with an adaptive quantization step.

**TP1-P4.32 DATA EMBEDDING IN MPEG-1/AUDIO LAYER II COMPRESSED DOMAIN USING SIDE INFORMATION**

*Akihiro Matsuoka, Kiyoshi Tanaka, Shinshu University, Japan; Akio Yoneyama, Yasuyuki Nakajima, KDDI R&D Laboratories, Inc., Japan*

In this work, we propose a data embedding scheme in MPEG-1/Audio Layer II compressed domain. Data embedding is conducted every AAU by using side information (location of sub-band allocated audio signal) as a data carrier. In general, non-zero signals concentrates in low and middle frequency bands. Therefore we utilize sub-bands that are not allocated audio signal in high frequency bands to embed information. The proposed scheme can increase payload while achieving rewritable (reversible) data embedding by choosing appropriate parameter. We verify the basic performance of our scheme through computer simulation by using some voice and music signals.

**TP1-P4.33 AN INNOVATIVE PIXEL SCORING METHOD FOR WATERMARKING OF BINARY DOCUMENT IMAGES**

*Edin Muharemagic, Savvy IT Solutions, United States; Borko Furht, Florida Atlantic University, United States*

In order to embed a watermark into a binary document image, some subset of image pixels needs to be modified. This modification will cause a document image distortion. Careful selection of image pixels can make distortion appear less visible. We propose a new binary document image pixel scoring method, the Structural Distortion Measure, whose objective is to identify image pixels whose modification, as part of a watermark embedding process, will minimize document image visible distortion.

**TP1-P4.34 WAVELET-BASED COLOR IMAGE WATERMARKING USING ADAPTIVE ENTROPY CASTING**

*Ming-Shing Hsieh, Aletheia University, Taiwan; Din-Chang Tseng, National Central University, Taiwan*

An adaptive robust image watermarking technique for color image authentication is proposed. In the proposed approach, the Y channel of a Yuv color host image and a concatenated RGB color-image watermark are decomposed into wavelet coefficients. The contextual entropies of the host wavelet coefficients are computed by a log-sum entropy measurement. Then the watermark wavelet coefficients are embedded into the host wavelet coefficients with larger entropy in the corresponding wavelet subbands. An adaptive casting strategy is utilized to embed the watermark coefficients for fully controlling the imperceptibility of watermarked images and the robustness of watermarks. The experimental results show that the proposed approach provides extra robustness against JPEG-compression and image-processing attacks; moreover, the proposed approach has no need of the original host image to extract watermarks.

**TP1-P4.35 SPATIALLY ADAPTIVE WAVELET THRESHOLDING FOR IMAGE WATERMARKING**

*Mahmood Al-khassaweneh, Selin Aviyente, Michigan State University, United States*

In this paper, we introduce a new robust image watermarking technique based on the Discrete Wavelet Transform (DWT). The proposed method extends the concept of image denoising to watermarking. A spatially adaptive wavelet thresholding method is used to select the coefficients to be watermarked. A multi-bit watermark is embedded into the discrete wavelet coefficients of the host image. A semi-blind watermark extraction algorithm is presented and the threshold for a given probability of false alarm is derived. The simulation results show that the proposed method outperforms a well-known DWT based watermarking method under most attacks including JPEG compression.

## **TP1-P4.36 A GEOMETRICALLY ROBUST WATERMARKING SCHEME BASED ON SELF-RECOGNITION WATERMARK PATTERN**

*Hefei Ling, Zhengding Lu, Fuhao Zou, Wugang Yuan, Huazhong University of Science and Technology, China*

Presently, almost all of the watermarking algorithms resistant to geometric attacks are dependent on the synchronization. In this paper, a geometrically robust watermarking scheme, which is not sensitive to synchronization, is proposed based on self-recognition watermark pattern. The structure of watermark pattern is elaborately designed, which makes the pattern recognized by its structure. Different watermark corresponds to different watermark pattern. The watermark bits can be decoded by recognizing the structures of detected pattern. This method's complexity is very low because no additional work is needed to meet the synchronization requirement and no area partitioning and normalization are needed. Experimental results show that this scheme is not only robust against geometric attacks including rotation, scaling, cropping, aspect ratio changing etc., but also resist numeric processes including Gaussian additive noise, low-pass filtering and lossy compression etc.

### **TP1-P5: Tuesday, July 11, 13:30 - 14:30**

#### **TP1-P5 Applications to Sports Video (Poster)**

Time: Tuesday, July 11, 13:30 - 14:30

Place: Toronto III

Chair: Ajay Divakaran, Mitsubishi Electric Research Labs

#### **TP1-P5.37 EXTRACTING STORY UNITS IN SPORTS VIDEO BASED ON UNSUPERVISED VIDEO SCENE CLUSTERING**

*Chunxi Liu, Qingming Huang, Shuqiang Jiang, Chinese Academy of Sciences, China; Weigang Zhang, Harbin Institute of Technology, China*

Many sports videos such as archery, diving and tennis have repetitive structure patterns. They are reliable clues to generate highlights, summarization and automatic annotation. In this paper, we present a novel approach to analyze these structure patterns in sports video to extract story units. First, an unsupervised scene clustering method for sports video is adopted to automatically categorize the video shots into several disparate scenes. Then, the clustering results are modeled by a transition matrix. Finally, the key scene shots are detected to analyze the structure patterns and extract the story units. Experimental results on several types of broad-cast sports video demonstrate that our approach is effective.

#### **TP1-P5.38 AUTOMATIC CLASSIFICATION OF FIELD OF VIEW IN VIDEO**

*Maria Zapata Ferrer, Mauro Barbieri, Hans Weda, Philips Research Laboratories, Netherlands*

Automatic systems are needed for audiovisual databases to efficiently index, browse, summarize and retrieve, because the amount of stored data is increasing tremendously. Historically film production techniques, have developed, in part, to convey e.g. meaning or atmosphere to the viewer. By studying these techniques, established guidelines for conveying meaning may be incorporated into automated tools for video analysis. In the current paper we present an approach in this area to classify different shot types, such as long shots, medium shots and close ups, which are important elements of video production. Based on a set of features calculated from the audiovisual content (e.g. presence of camera motion and size of detected faces), a Bayesian classifier distinguishes between six different shot types. The performance of this novel generic field of view classifier in terms of precision and recall is promising.

#### **TP1-P5.39 HIGHLIGHT SUMMARIZATION IN SPORTS VIDEO BASED ON REPLAY DETECTION**

*Zhao Zhao, ShuQiang Jiang, Qingming Huang, Chinese Academy of Sciences, China; GuangYu Zhu, Harbin Institute of Technology, China*

Highlight summarization technology has been studied widely in sports video analysis. In this paper, we propose a highlight summarization system based on replays. First the replay clips in the sports video are extracted as the highlight candidates. Then the features including audio energy and motion activity are employed to rank the arousal level of the replay clips. Finally we model the highlight with the arousal rank to generate summarization. The contribution of this paper concentrates on two aspects. Firstly, Event-Replay (ER) structure is proposed and some new features are employed to represent the arousal levels of ER for general sports video. Secondly a novel highlight model is proposed considering the inter-relation of ERs. The experiments evaluate the rationality of the system.

#### **TP1-P5.40 EVENT-IMPORTANCE BASED CUSTOMIZED AND AUTOMATIC CRICKET HIGHLIGHT GENERATION**

*Maheshkumar Kolekar, Somnath Sengupta, Indian Institute of Technology, Kharagpur, India*

In this paper, we present a novel approach towards customized and automated generation of sports highlights from its extracted events and semantic concepts. A recorded sports video is first divided into slots, based on the game progress and for each slot, an importance-based concept and event selection is proposed to include those in the highlights. Using our approach, we have successfully extracted highlights from recorded video of cricket match.

#### **TP1-P5.41 SPORTS PROGRAM BOUNDARY DETECTION**

*Regunathan Radhakrishnan, Ajay Divakaran, Mitsubishi Electric Research Laboratories, United States; Isao Otsuka, Mitsubishi Electric Corporation, Japan*

In the recent years, consumer devices that can record broadcast video have become prevalent. Such devices rely on the program guide information about a program's start time and end time for recording. However, sports broadcasts can run over the specified time sometimes. In this paper, we propose a framework based on an audio classification framework to correctly detect the end of a sports broadcast so as to enable complete recording of the game. Our experimental results show that the proposed algorithm can detect sports program boundaries with a high accuracy.

#### **TP1-P5.42 SOCCER HIGHLIGHT DETECTION USING TWO-DEPENDENCE BAYESIAN NETWORK**

*Jianguo Li, Tao Wang, Wei Hu, Mingliang Sun, Yiming Zhang, Intel China Research Center, China*

Soccer highlight detection is an active research topic in recent years. One of the difficult problems is how to effectively fuse multimodal cues, i.e. audio, visual and textual information, to improve the detection performance. This paper proposes a novel two-dependence Bayesian network (2d-BN) based fusion approach to soccer highlight detection. 2d-BN is a particular Bayesian network which assumes that each variable depends on two other variables at most. Through this assumption, 2d-BN can not only characterize the relationships among features but also be trained efficiently. Extensive experiments demonstrate the effectiveness of the proposed method.

### **TP2-L1: Tuesday, July 11, 14:30 - 15:50**

#### **TP2-L1 Object Recognition (Lecture)**

Time: Tuesday, July 11, 14:30 - 15:50

Place: Carmichael

Chair: Yong Rui, Microsoft

14:30 - 14:50

#### **TP2-L1.1 AUTOMATIC MULTI-PLAYER DETECTION AND TRACKING IN BROADCAST SPORTS VIDEO USING SUPPORT VECTOR MACHINE AND PARTICLE FILTER**

*Guangyu Zhu, Harbin Institute of Technology, China; Changsheng Xu, Institute for Infocomm Research, Singapore; Qingming Huang, Chinese Academy of Sciences, China; Wen Gao, Harbin Institute of Technology, China*

In this paper, a novel multiple objects detection and tracking approach based on support vector machine and particle filter is proposed to track players in broadcast sports video. Compared with previous work, the contributions of this paper are focused on three aspects. First, an improved particle filter called SVR particle filter is proposed as the player tracker by integrating support vector regression (SVR) into sequential Monte Carlo framework. SVR particle filter enhances the performance of classical particle filter with small sample set and improves the efficiency of tracking system. Second, support vector classification combined with playfield segmentation is employed to automatically detect the players in sports video as the initialization of tracker. Third, a unified framework for automatic object detection and tracking is proposed based on support vector machine and particle filter. The experimental results are encouraging and demonstrate that our approach is effective.



14:50 - 15:10

**TP2-L1.2    RECOGNIZING FACES IN RECORDED MEETINGS VIA MRC-BOOSTING**

*Xun Xu, University of Illinois at Urbana-Champaign, United States; Yong Rui, Microsoft Research, United States; Thomas S. Huang, University of Illinois at Urbana-Champaign, United States*

Person-based indices and timelines can enable fast and non-linear access to recorded meetings. This paper focuses on how to automatically construct those indices and timelines by using face recognition techniques. While there exist extensive research in generic face recognition, recognizing faces in recorded meetings is still an understudied area. Real-world meeting videos impose several interesting and unique challenges including complex lighting, low imaging quality, and large variations in head pose and size. In this paper, a promising approach based on MRC-Boosting is presented to address these challenges, which achieves encouraging performance on real-world meeting videos and shows superior accuracy and robustness compared to two popular existing approaches.

15:10 - 15:30

**TP2-L1.3    SELECTING KERNEL EIGENFACES FOR FACE RECOGNITION WITH ONE TRAINING SAMPLE PER SUBJECT**

*Jie Wang, Konstantinos N. Plataniotis, Anastasios (Tas) N. Venetsanopoulos, University of Toronto, Canada*

It is well-known that supervised learning techniques such as linear discriminant analysis (LDA) often suffer from the so called small sample size problem when apply to solve face recognition problems. This is due to the fact that in most cases, the number of training samples is much smaller than the dimensionality of the sample space. The problem becomes even more severe if only one training sample is available for each subject. In this paper, followed by the well-known unsupervised technique, kernel principal component analysis (KPCA), a novel feature selection scheme is proposed to establish a discriminant feature subspace in which the class separability is maximized. Extensive experiments performed on the FERET database indicate that the proposed scheme significantly boosts the recognition performance of the traditional KPCA solution.

15:30 - 15:50

**TP2-L1.4    PARTIAL LDA VS PARTIAL PCA**

*Antonio Rama, Francesc Tarres, Technical University of Catalonia (UPC), Spain*

Recently, 3D face recognition algorithms have outperformed 2D conventional approaches by adding depth data to the problem. However, independently of the nature (2D or 3D) of the approach, the majority of them required the same data format in the test stage than the data used for training the system. This issue represents the main drawback of 3D face research since 3D data should be acquired under highly controlled conditions and in most of the cases with the collaboration of the subject to be recognized. Thus, in real world applications (control access points or surveillance) this kind of 3D data may not be available during the recognition process. This should follow to a new paradigm of using some mixed 2D-3D face recognition systems where 3D data is used in the training but either 2D or 3D information can be used in the recognition depending on the scenario. Following this new concept, Partial Linear Discriminant Analysis (PLDA) is presented in this paper. Preliminary results have shown an improvement with respect to the Partial PCA approach [1].

**TP2-L2      Multimedia Interfaces: Speech and Audio (Lecture)**

Time:        Tuesday, July 11, 14:30 - 15:50

Place:       Johnston

Chair:       Susanto Rahardja, Institute for Infocomm Research

14:30 - 14:50

**TP2-L2.1    A NOVEL INTERFACE FOR AUDIO SEARCH**

*Sarah Ali, Parham Aarabi, University of Toronto, Canada*

In this paper a novel cyclic interface for searching through a song database is proposed. The method, which merges multiple audio streams on a server and broadcasts only a single merged stream, allows the user to hear different parts of each audio stream by cycling through all available streams. Experimental results on 21 users illustrate that the proposed interface requires less listening time as compared to traditional list-based interfaces when the desired song/audio clip is among one of the audio streams. The average search time for the proposed interface was 7.3 seconds, compared to 12.1 seconds for the traditional list-based interface when searching for a song which is included among the audio streams.

14:50 - 15:10

**TP2-L2.2    REMOTE VOICE ACQUISITION IN MULTIMODAL SURVEILLANCE**

*Weihong Li, City University of New York - Graduate Center, United States; Zhigang Zhu, George Wolberg, City University of New York - City College, United States*

Multimodal surveillance systems using visible/IR cameras and other sensors are widely deployed today for security purpose, particularly when subjects are at a large distance. However, audio information as an important data source has not been well explored. One of the reasons is because audio detection using microphones needs installation close to the subjects in monitoring. In this paper, we investigate a novel “optical” sensor, called Laser Doppler Vibrometer (LDV), for capturing voice signals in a very large range to realize a truly remote and multimodal surveillance system. Speech enhancement approaches are studied based on the characteristics of LDV Audio. Experimental results show that remote voice detection via an LDV is promising when choosing appropriate targets close to human subjects in the environment.

15:10 - 15:30

**TP2-L2.3    EMOTION RECOGNITION FROM NOISY SPEECH**

*Mingyu You, Chun Chen, Jiajun Bu, Jia Liu, Zhejiang University, China; Jianhua Tao, Chinese Academy of Sciences, China*

This paper presents an emotion recognition system from clean and noisy speech. Geodesic distance was adopted to preserve the intrinsic geometry of emotional speech. Based on the geodesic distance estimation, an enhanced Lipschitz embedding was developed to embed the 64-dimensional acoustic features into a six-dimensional space. In order to avoid the problems brought by noise reduction, emotion recognition from noisy speech was performed directly. Linear Discriminant Analysis (LDA), Principal Component Analysis (PCA) and feature selection by Sequential Forward Selection (SFS) with Support Vector Machine (SVM) were also included to compress acoustic features before classifying the emotional states of clean and noisy speech. Experimental results demonstrate that compared with other methods, the proposed system makes approximately 10% improvement. The performance of our system is also robust when speech data is corrupted by increasing noise.

**TP2-L3 Coding Algorithms and Standards II (Lecture)**

Time: Tuesday, July 11, 14:30 - 15:50

Place: Thomson

Chair: Liang Zhang, Communications Research Centre

14:30 - 14:50

**TP2-L3.1 CONSTANT QUALITY AIMED BIT ALLOCATION FOR 3D WAVELET BASED VIDEO CODING**

*Zefeng Ni, Jianfei Cai, Nanyang Technological University, Singapore*

MCTF has been widely used in wavelet based video coding. For MCTF based codecs, a fundamental question is how to allocate bits to each temporal band so that certain degree of constant quality can be achieved. In this paper, we propose a novel approach for constant quality aimed bit allocation among T-bands for adaptive stored video streaming. The basic idea of our proposed scheme is to adjust the energy gains based on an empirical model to compensate the different contributions from different types of T-bands, and more or less equally distribute the distortions among the T-bands at the same level. Experimental results show that our proposed bit allocation can greatly reduce the PSNR fluctuation with only slight degradation in average PSNR.

14:50 - 15:10

**TP2-L3.2 EMBEDDED CODING OF THE MOTION-COMPENSATED 3-D WAVELET COEFFICIENTS BY CONCATENATING SPATIAL AND TEMPORAL ORIENTATION TREES**

*Liang Zhang, Communications Research Centre Canada, China*

This paper focuses on the issue of coding of the spatiotemporal 3-D wavelet coefficients. Experimental examination found that there are temporal dependencies among the wavelet coefficients at different temporal decomposition levels. Temporal orientation trees are introduced to exploit these dependencies to improve the coding efficiency. Statistical properties of the wavelet coefficients are measured using test video sequences. 3-D orientation trees that are built by spatial orientation trees followed by temporal orientation trees are exploited to magnitude-order the spatiotemporal 3-D wavelet coefficients. The experimental results confirmed that the coder with temporal orientation trees outperformed the coder without them. The coding gain could be up to 0.5 dB.

15:10 - 15:30

**TP2-L3.3 BIT-STREAM SWITCHING IN MULTIPLE BIT-RATE VIDEO STREAMING USING WYNER-ZIV CODING**

*Wen Sun, University of Science and Technology of China, China; Yan Lu, Feng Wu, Microsoft Research Asia, China*

It has been commonly recognized that multiple bit-rate (MBR) encoding provides a concise method for video streaming over bandwidth-fluctuant networks. The key problem of the MBR technique lies in how to seamlessly switch one bit-stream to another one. To tackle this problem, we propose a bit-stream switching framework based on the Wyner-Ziv coding. Within the propose framework, the multiple bit-streams can be individually encoded without data exchange, which also supports the random switching at any desired frame without affecting the original coding efficiency of the regular bit-stream. In particular, two different implementation schemes under the same framework are presented. Different from the traditional switching schemes, the proposed method can use the same switching frame for the switching from any other bit-stream to the current one, which means less storage and less encoding efforts. Simulation results and comparison between the proposed method and the traditional switching method in H.264 are also presented.

15:30 - 15:50

**TP2-L3.4 A NOVEL RESYNCHRONIZATION METHOD FOR SCALABLE VIDEO OVER WIRELESS CHANNEL**

*Yu Wang, Lap-Pui Chau, Kim-Hui Yap, Nanyang Technological University, Singapore*

A scalable video coder generates scalable compressed bit-stream, which can provide different types of scalability depend on different requirements. This paper proposes a novel resynchronization method for the scalable video with combined temporal and quality (SNR) scalability. The main purpose is to improve the robustness of the transmitted video. In the proposed scheme, the video is encoded into scalable compressed bit-stream with combined temporal and quality scalability. The significance of each enhancement layer unit is estimated properly. A novel resynchronization method is proposed where joint group of picture (GOP) level and picture level insertion of resynchronization marker approach is applied to insert different amount of resynchronization markers in different enhancement layer units for reliable transmission of the video over error-prone channels. It is demonstrated from the experimental results that the proposed method can perform graceful degradation under a variety of error conditions and the improvement can be up to 1 dB compared with the conventional method.

**TP2-L4: Tuesday, July 11, 14:30 - 15:50**

**TP2-L4 P2P Multimedia IV (Special Session)**

Time: Tuesday, July 11, 14:30 - 15:50

Place: Toronto I

Co-Chairs: Thinh Nguyen, Oregon State University and Heather Yu, Panasonic

14:30 - 14:50

**TP2-L4.1 EFFICIENT VIDEO DISSEMINATION IN STRUCTURED HYBRID P2P NETWORKS**

*Thinh Nguyen, Krishnan Kolazhi, Rohit Kamath, Oregon State University, United States; Sen-ching Cheung, University of Kentucky, United States*

In this paper, we propose a structured hybrid P2P mesh for optimal video dissemination from a single source node to multiple receivers in a bandwidth-asymmetric network such as Digital Subscriber Line (DSL) access network. Our hybrid P2P structured mesh consists of one or more supernodes and a large number of streaming nodes connected together in special manner designed for streaming and real-time video dissemination. The supernode is responsible for managing node's joining and leaving, while the streaming nodes are responsible for actual data transmissions among peers. Our proposed hybrid P2P structured mesh is designed to achieve scalability, bandwidth fairness, low delay, and high video quality. Our experimental Internet-wide system consisting of PlanetLab nodes demonstrates that it is highly scalable, and can provide low delay and high bandwidth video dissemination.

14:50 - 15:10

**TP2-L4.2 INVOLVING CLIENTS IN THE DISTRIBUTION OF VIDEOS ON DEMAND**

*Santosh Kulkarni, Jehan-François Pâris, University of Houston, United States*

We present a stream tapping protocol that involves clients in the video distribution process. As in conventional stream tapping, our protocol lets new clients tap the most recent broadcast of the video they are watching. While conventional stream tapping required the server to send to these clients the part of the video they missed, our protocol delegates this task to the clients that are already watching the video, thus greatly reducing the workload of the server. Unlike previous solutions involving clients in the video distribution process, our protocol works with clients that can only upload video data at fraction of the video consumption rate.

15:10 - 15:30

**TP2-L4.3 PATH-DIVERSITY OVERLAY RETRANSMISSION ARCHITECTURE FOR RELIABLE MULTICAST**

*Wenjun Zeng, Yingnan Zhu, Haibin Lu, Hongbing Jiang, University of Missouri - Columbia, United States*

IP-multicast is a bandwidth efficient transmission mechanism for group communications. Reliability in IP-multicast, however, poses a set of significant challenges. To address the reliability and scalability issues in IP-multicast, this paper proposes a novel overlay retransmission architecture that exploits path-diversity by taking advantages of both IP multicast and an overlay network. We show that the proposed path diversity overlay retransmission architecture has the potential to significantly improve the reliability, delay, playback quality, and scalability of IP-multicast based multimedia applications. The general concept of using P2P overlay networks to help improve the QoS performance of multimedia applications as illustrated in this paper is expected to have significant impact on the deployment of next generation multimedia services.

15:30 - 15:50

**TP2-L4.4 USING P2P NETWORKS FOR ERROR RECOVERY IN MBMS APPLICATIONS**

*Praveen Sanigepalli, Motorola, United States; Hari Kalva, Borko Furht, Florida Atlantic University, United States*

The wireless networks are notoriously error prone and all errors cannot be prevented in real-time communications. The problem of error correction becomes even more challenging in mobile multicast/broadcast applications. The mobile devices are being equipped with multiple modems that could work simultaneously; for example, devices with both GSM (WAN) and WLAN networks such as WiFi. These multi-modal devices can use the second network to improve their error resilience. We propose a P2P approach to establish and utilize an error recovery channel on a secondary network for multi-user video applications. The mobile devices within the vicinity can utilize the WLAN network to form a P2P network for error recovery purpose. We developed and evaluated three error recovery models for error recovery over secondary networks. The proposed models balance the response time, bandwidth utilization, fairness, and unnecessary data received.

**TP2-L5: Tuesday, July 11, 14:30 - 15:50**

**TP2-L5 3D-TV: Primed for Success? I (Special Session)**

Time: Tuesday, July 11, 14:30 - 15:50

Place: Toronto II

Co-Chairs: Liang Zhang, Communications Research Centre Canada, Wa James Tam, Communications Research Centre Canada and Hirokazu Yamanoue, NHK Science & Technical Research Laboratories

14:30 - 14:50

**TP2-L5.1 CHALLENGES AND OPPORTUNITIES IN VIDEO CODING FOR 3D TV**

*Hari Kalva, Lakis Christodoulou, Liam Mayron, Oge Marques, Borko Furht, Florida Atlantic University, United States*

This paper explores the challenges opportunities in developing and deploying 3D TV services. The 3D TV services can be seen as a general case of the multi-view video that has been receiving significant attention lately. The keys to a successful 3D TV experience are the availability of content, the ease of use, the quality of experience, and the cost of deployment. Recent technological advances have made possible experimental systems that can be used to evaluate the 3D TV services. We have developed a 3D TV prototype and have currently conducting our first user study to evaluate the quality and experience. These experiences have allowed us to identify challenges and opportunities in developing 3D TV services.

14:50 - 15:10

**TP2-L5.2 A ROADMAP FOR AUTOSTEREOSCOPIC MULTI-VIEWER DOMESTIC TV DISPLAYS**

*Phil Surman, De Montfort University, United Kingdom; Klaus Hopf, Fraunhofer – Heinrich Hertz Institute, Germany; Ian Sexton, Wing Lee, Richard Bates, De Montfort University, United Kingdom*

This paper presents a brief overview of the current technologies and technical approaches that may lead to viable and user-acceptable domestic autostereoscopic multi-viewer television displays. It illustrates the performance attributes of the various technological approaches and points to the most likely approaches to succeed within the next 10 years. Finally, it shows possible timescales for the enabling technologies for 3D display, and concludes that multi-user autostereoscopic displays may be the first to gain widespread use.

15:10 - 15:30

**TP2-L5.3 HUMAN FACTORS FOR STEREOSCOPIC IMAGES**

*Kazuhiko Ukai, Waseda University, Japan*

Human factors of stereoscopic images, the background, theories, and observations on visual stress caused by viewing stereoscopic motion images are introduced. After studies on accommodation and convergence are surveyed and an explanation regarding the characteristics of these functions is offered, difficulties in fusing two images, fatigue possibly caused by the discrepancy between accommodative and convergence stimuli that are included in the image are discussed. Changes in oculomotor function after viewing stereoscopic images, and irreversible changes in oculomotor functions are discussed.

15:30 - 15:50

**TP2-L5.4 THE DIFFERENCES BETWEEN TOED-IN CAMERA CONFIGURATIONS AND PARALLEL CAMERA CONFIGURATIONS IN SHOOTING STEREOSCOPIC IMAGES**

*Hirokazu Yamanoue, NHK, Japan*

A fundamental element of stereoscopic image production is to geometrically analyze the conversion from real space to stereoscopic images by binocular parallax under various shooting and viewing conditions. This paper reports on this analysis, particularly on the setting of the optical axes of 3D cameras, which has received little attention in the past. The parallel camera configuration maintains linearity during the conversion from real space to stereoscopic images. Under the camera configuration, optical axes are kept parallel to each other and apparent magnification (lateral magnification) of a shooting target is not dependent on the shooting distance. But the toed-in camera configuration often can not maintain linearity during the conversion from real space to stereoscopic images where the apparent magnification of a shooting target is dependent on the shooting distance.

**TP2-L6: Tuesday, July 11, 14:30 - 15:50**

**TP2-L6 H.264 I (Lecture)**

Time: Tuesday, July 11, 14:30 - 15:50

Place: Varley

Chair: Thomas Wiegand, Fraunhofer HHI

14:30 - 14:50

**TP2-L6.1 AN EFFICIENT CRITERION FOR MODE DECISION IN H.264/AVC**

*Yu-Kuang Tu, Jar-Ferr Yang, National Cheng Kung University, Taiwan; Ming-Ting Sun, University of Washington, United States*

In this paper, an efficient cost function for mode decision in H.264/AVC is proposed. The proposed cost function is based on integer transform coefficients, where the rate and the distortion are jointly modeled by the number of nonzero quantized coefficients, the sum of absolute integer transformed differences (SAITD) and sum of squared integer transformed differences (SSITD). Comparing to the high-complexity cost function, which should be calculated from real bit-consumption and true reconstructed distortion for each coding mode, the proposed efficient cost function can achieve 79.93% and 22.61% time savings of computing rate-distortion cost and overall encoding, respectively, while introducing only slight degradation with 1.05% bit-rate increment and 0.049dB PSNR drop.

14:50 - 15:10

**TP2-L6.2 A FAST MODE SELECTION ALGORITHM IN H.264 VIDEO CODING**

*Donghyung Kim, Jechang Jeong, Hanyang University, Republic of Korea*

For improvement of coding efficiency, the H.264 video coding standard uses new coding tools, such as variable block size, quarter-pixel-accuracy motion estimation, multiple reference frames, intra prediction, loop filter, etc. Using these coding tools, H.264 achieves significant improvement in coding efficiency compared with existing standards. However, encoder complexity increases tremendously. Among the tools, the macroblock mode selection and the motion estimation contribute most to total encoder complexity. This paper focuses on complexity reduction in macroblock mode selection. Of all macroblock modes which can be selected, inter8×8 and intra4×4 have the highest complexity. We propose two methods for complexity reduction of inter8×8 and intra4×4 by using the costs of the other macroblock modes. Simulation results show that the proposed methods save up to 57.7% of total encoding time compared with the H.264 reference implementation, whereas the average PSNR only decreases less than 0.05dB.

15:10 - 15:30

**TP2-L6.3 A MOTION VECTOR RECOVERY ALGORITHM FOR TEMPORAL ERROR CONCEALMENT USING OPTICAL FLOW IN H.264 VIDEO CODING**

*Donghyung Kim, Sanghyup Cho, Jechang Jeong, Hanyang University, Republic of Korea*

For the improvement of coding efficiency, the H.264 standard uses new coding tools. Among the tools, motion estimation using smaller block sizes leads to higher correlation between the motion vectors of neighboring blocks. This characteristic of H.264 is useful for motion vector recovery to conceal a lost macroblock. In this paper, we propose the motion vector recovery method based on optical flow in H.264. We first determine optical flow region to alleviate the complexity, and choose initial value of flow velocity using neighboring motion vectors of a lost macroblock. And then the proposed method recovers the motion vectors of 4×4 blocks included in a lost macroblock using the weighted average of obtained flow velocities. Simulation results show that our proposed method gives higher objective and subjective visual qualities than conventional approaches.

15:30 - 15:50

**TP2-L6.4 EFFICIENT COMPRESSION OF MULTI-VIEW VIDEO EXPLOITING INTER-VIEW DEPENDENCIES BASED ON H.264/MPEG4-AVC**

*Philipp Merkle, Karsten Müller, Aljoscha Smolic, Thomas Wiegand, Fraunhofer – Heinrich Hertz Institute, Germany*

Efficient Multi-view coding requires coding algorithms that exploit temporal, as well as inter-view dependencies between adjacent cameras. Based on a spatiotemporal analysis on the multi-view data set, we present a coding scheme utilizing an H.264/MPEG4-AVC codec. To handle the specific requirements of multi-view datasets, namely temporal and inter-view correlation, two main features of the coder are used: hierarchical B pictures for temporal dependencies and an adapted prediction scheme to exploit inter-view dependencies. Both features are set up in the H.264/MPEG4-AVC configuration file, such that coding and decoding is purely based on standardized software. Additionally, picture reordering before coding to optimize coding efficiency and in-verse reordering after decoding to obtain individual views are applied. Finally, coding results are shown for the proposed multi-view coder and compared to simulcast anchor and simulcast hierarchical B picture coding.



**TP2-P1 Feature Extraction and Classification II (Poster)**

Time: Tuesday, July 11, 14:50 - 15:50

Place: Toronto III

Chair: Pascal Frossard, Swiss Federal Institute of Technology (EPFL)

**TP2-P1.1 MULTISCALE EDGE-BASED TEXT EXTRACTION FROM COMPLEX IMAGES**

*Xiaoqing Liu, Jagath Samarabandu, University of Western Ontario, Canada*

Text that appears in images contains important and useful information. Detection and extraction of text in images have been used in many applications. In this paper, we propose a multiscale edge-based text extraction algorithm, which can automatically detect and extract text in complex images. The proposed method is a general-purpose text detection and extraction algorithm, which can deal not only with printed document images but also with scene text. It is robust with respect to the font size, style, color, orientation, and alignment of text and can be used in a large variety of application fields, such as mobile robot navigation, vehicle license detection and recognition, object identification, document retrieving, page segmentation, etc.

**TP2-P1.2 NEW AREA MATRIX-BASED AFFINE-INVARIANT SHAPE FEATURES AND SIMILARITY METRICS**

*Carlos Ramon Pantaleon Dionisio, Hae Yong Kim, Universidade de São Paulo, Brazil*

A near-planar object seen from different viewpoints results in differently deformed images. Under some assumptions, viewpoint changes can be modeled by affine transformations. Shape features that are affine-invariant (af-in) must remain constant with the changes of the viewpoint. Similarly, shape similarity metrics that are af-in must rate the difference between two shapes, regardless of their viewpoints. Af-in shape features and similarity metrics can be used for the shape classification and retrieval. In this paper, we propose a new set of af-in shape features and similarity metrics. They are based on the area matrix, a structure that contains multiscale information about the shape. Experimental results indicate that the proposed techniques are robust to viewpoint changes and can rate correctly the dissimilarities between the shapes.

**TP2-P1.3 A ROBUST ENTROPY-BASED AUDIO-FINGERPRINT**

*Antonio C. Ibarrola, Edgar Chávez, Universidad Michacana de San Nicolás de Hidalgo, Mexico*

Audio Fingerprints (AFP's) are compact, content-based representations of audio signals used to measure distances among them. An AFP has to be small, fast computed and robust to signal degradations. In this paper an entropy based AFP is presented that performed very well when the signal was corrupted with lossy compression, scaling and even 1KHz Low-pass filtering in the experiments. The AFP is determined by computing the instantaneous amount of information of the audio signal in two-second frames with fifty percent overlapping, the resulting entropy signal is binary coded in order to compare different interpretations (e.g. live vs. studio recording) of the same song with good results. The AFP's robustness is compared with that of Haitsma-Kalker's Hash string based AFP with encouraging results

**TP2-P1.4 FACE RECOGNITION USING 3D SUMMATION INVARIANT FEATURES**

*Wei-Yang Lin, Kin-Chung Wong, Yu Hen Hu, Nigel Boston, University of Wisconsin - Madison, United States*

In this paper, we developed a family of 2D and 3D invariant features with applications to 3D human faces recognition. The main contributions of this paper are: (a) systematically deriving a family of novel features, called {nit summation invariant} that are invariant to Euclidean transformation in both 2D and 3D; (b) developing an effective method to apply summation invariant to the 3D face recognition problem. Tested with the 3D data from the Face Recognition Grand Challenge v1.0 dataset, the proposed new features exhibit achieves a performance that rivals the best 3D face recognition algorithms reported so far.

**TP2-P1.5 A NEW APPROACH FOR TEXTURE FEATURES EXTRACTION: APPLICATION FOR TEXT LOCALIZATION IN VIDEO IMAGES**

*Bassem Bouaziz, Walid Mahdi, MIRACL, Tunisia; Mohsen Ardabilain, LIRIS, France; Abdelmajid Ben Hamadou, MIRACL, Tunisia*

In this paper we present a new texture feature extraction approach. Existing methods are generally time consuming and sensible to image complexity in terms of texture's regularity, directionality and coarseness. So that we propose a method which provide both rapidity and accuracy to extract and characterize texture features. It's based on Hough Transform technique combined with an extremity segment's neighbourhood analysis and a new computation algorithm to extract segments and detect regularity. Experimental results show that this approach is robust and can be applied not only to texture analysis but also to detect text within video images.

#### **TP2-P1.6 CAMERA MOTION DETECTION USING VIDEO MOSAICING**

*Masaki Naito, Kazunori Matsumoto, Keiichiro Hoashi, Fumiaki Sugaya, KDDI R&D Laboratories, Inc, Japan*

In this paper, camera motion detection methods using a background image generated by video mosaicing based on the correlation between feature points on a frame pair are described. In this method, a telop reduction method, iterative foreground and background image separation method and appropriate frame pair selection from consecutive frames are introduced to generate background images accurately. Parameters indicating the location of each frame on the background image are retrieved and used to detect the camera motion. Except for the simple threshold-based method, a method using Hidden Markov models (HMMs) is introduced to detect variable length camera motion based on the maximum likelihood criterion. The effectiveness of the proposed method is evaluated by using a TRECVID 2005 low-level feature extraction task.

#### **TP2-P1.7 A GLOBAL CORRESPONDENCE FOR SCALE INVARIANT MATCHING USING MUTUAL INFORMATION AND THE GRAPH SEARCH**

*Hyun-Ho Jeon, Andrea Basso, Peter Driessen, University of Victoria, Canada*

In this paper we propose a novel approach to find a global correspondence between two images by maximizing mutual information in the presence of large scale changes and rotations. Our experimental results show that the proposed local descriptor and graph-based search allows robust point matching.

#### **TP2-P1.8 SELF-SUPERVISED LEARNING FOR ROBUST VIDEO INDEXING**

*Ralph Ewerth, Bernd Freisleben, University of Marburg, Germany*

The performance of video analysis and indexing algorithms strongly depends on the type, content and recording characteristics of the analyzed video. Current video indexing approaches often make use of thresholding techniques or supervised learning which requires labeling of possibly large training sets. Furthermore, the application of the same training model or parameters might lead to a sub-optimal indexing accuracy for a given video. In this paper, we propose to use a novel self-supervised learning framework for robust video indexing to address this issue. Based on an initial classification result for a given video, the best features are selected by Adaboost and are then used to train SVM (support vector machine) classifiers, all on the given video. Finally, a specialized ensemble of classifiers is employed for the given video for decision making. Experimental results show that a state-of-the-art video cut detection approach can be significantly improved by the self-supervised learning approach.

#### **TP2-P1.9 DISTRIBUTED SVM APPLIED TO IMAGE CLASSIFICATION**

*Effrosyni Kokiopoulou, Pascal Frossard, Ecole Polytechnique Fédérale de Lausanne (EPFL), Switzerland*

This paper proposes an algorithm for distributed classification, based on a SVM scheme. The contribution of each support vector is approximated by low complexity distributed thresholding over sub-dictionaries, whose union forms a redundant dictionary of atoms that spans the space of the observed signal. Redundant dictionaries allow for sparse representation of the observed signal, hence a good approximation of the support vector contributions, which is moreover robust to noise. The algorithm is applied to distributed image classification, in the context of handwritten digit recognition in a sensor network. The experimental results indicate that the proposed method is capable of achieving the same classification performance as the standard (non distributed) SVM, with an increased resiliency to noise.

#### **TP2-P1.10 PROBABILISTIC MULTIMODALITY FUSION FOR EVENT BASED HOME PHOTO CLUSTERING**

*Tao Mei, Bin Wang, University of Science and Technology of China, China; Xian-Sheng Hua, Microsoft Research Asia, China; He-Qin Zhou, University of Science and Technology of China, China; Shipeng Li, Microsoft Research Asia, China*

This paper presents a novel probabilistic approach to fusing multimodal metadata for event based home photo clustering. Photo events are characterized by the coherence of multimodality including time, content and camera settings. We incorporate these multimodal metadata into a unified probabilistic framework, in which event is taken as a latent semantic concept and discovered by fitting a generative model through an Expectation-Maximization (EM) algorithm. This approach is general and unsupervised, without any training procedure or predefined threshold. The experimental evaluations on 14k photos taken by 10 amateur photographers have indicated the effectiveness and efficiency of the proposed framework in browsing and searching personal photo collections.

**TP2-P2 Multimedia Database Query Processing and Retrieval (Poster)**

Time: Tuesday, July 11, 14:50 - 15:50

Place: Toronto III

Chair: Apostol Natsev, IBM - T.J. Watson Research Center

**TP2-P2.11 SEMANTIC MULTIMEDIA RETRIEVAL USING LEXICAL QUERY EXPANSION AND MODEL-BASED RERANKING**

*Alexander Haubold, Columbia University, United States; Apostol (Paul) Natsev, Milind R. Naphade, IBM T. J. Watson Research Center, United States*

We present methods for improving text search retrieval of visual multimedia content by applying a set of visual models of semantic concepts from a lexicon of concepts deemed relevant for the collection. Text search is performed via queries of words or fully qualified sentences, and results are returned in the form of ranked video clips. Our approach involves a query expansion stage, in which query terms are compared to the visual concepts for which we independently build classifier models. We leverage a synonym dictionary and WordNet similarities during expansion. Results over each query are aggregated across the expanded terms and ranked. We validate our approach on the TRECVID 2005 broadcast news data with 39 concepts specifically designed for this genre of video. We observe that concept models improve search results by nearly 50% after model-based re-ranking of text-only search. We also observe that purely model-based retrieval significantly outperforms text-based retrieval on non-named entity queries.

**TP2-P2.12 LOCALLY EMBEDDED LINEAR SUBSPACES FOR EFFICIENT VIDEO INDEXING AND RETRIEVAL**

*Zhu Li, Motorola Labs, United States; Li Gao, Aggelos K. Katsaggelos, Northwestern University, United States*

Abstract Efficient indexing is a key in content-based video retrieval solutions. In this paper we represent video sequences as traces via scaling and linear transformation of the frame luminance field. Then an appropriate lower dimensional subspace is identified for video trace indexing. We also develop a trace geometry matching algorithm for retrieval based on average projection distance with a locally embedded distance metric. Simulation results demonstrated the effectiveness of the proposed approach in both speed and retrieval accuracy.

**TP2-P2.13 IMAGE RETRIEVAL BASED ON USER-SPECIFIED FEATURES IN MULTI-CLUSTER QUERIES**

*Khanh Vu, Kien Hua, Soontharee Koopairojn, University of Central Florida, United States*

In a typical image retrieval system, all visual features of query images are used to determine image similarity. Thus, users are left to decide whether or not to include images that not only contain desirable features but also irrelevant ones. Fewer examples or a contaminated set of more could compromise the retrieval effectiveness of most similarity measures. In this paper, we extend our previous approach that allows users define queries by specifying relevant features present in image examples. The extended technique support queries decomposed in multiple clusters, each forming a subquery. Our experimental results have shown a remarkable improvement in retrieval performance.

**TP2-P2.14 USING IMPLICIT RELEVANCE FEEDBACK TO ADVANCE WEB IMAGE SEARCH**

*En Cheng, Huazhong University of Science and Technology, China; Feng Jing, Mingjing Li, Wei-Ying Ma, Microsoft Research Asia, China; Hai Jin, Huazhong University of Science and Technology, China*

Although relevance feedback has been extensively studied in content-based image retrieval in the academic area, no commercial search engine has employed the idea. There are several obstacles for Web image search engines in applying relevance feedback. To overcome these obstacles, we proposed an efficient implicit relevance feedback mechanism. The proposed mechanism shows advantage over traditional relevance feedback methods in the following three aspects. Firstly, instead of enforcing the users to make explicit judgment on the results, our method regards user's clickthrough data as implicit relevance feedback which release burden from users. Secondly, a hierarchical image search results clustering algorithm is proposed to semantically organize the search results. Using the clustering results as features, our relevance feedback scheme could catch and reflect users' search intention precisely. Lastly, unlike traditional relevance feedback user interface which hardly substitutes subsequent results for previous ones, our method employed friendly recommendation rather than substitution to let the user narrow down on the refined images. To evaluate the implicit relevance feedback mechanism, comprehensive user studies were performed.

**TP2-P2.15 REGION-BASED IMAGE RETRIEVAL USING RADIAL BASIS FUNCTION NETWORK**

*Kui Wu, Kim-Hui Yap, Lap-Pui Chau, Nanyang Technological University, Singapore*

This paper presents a new framework that integrates relevance feedback into region-based image retrieval (RBIR) systems based on radial basis function network (RBFN). A modified unsupervised subtractive clustering algorithm is proposed for RBFN center selection according to the characteristics of region-based image representation. A new kernel function of RBFN is introduced for image similarity comparison under region-based representation. The underlying network parameters (weight and width) are then optimized using a supervised gradient-descent training strategy. Experimental results using a database of 10,000 images demonstrate the effectiveness of the proposed hybrid learning approach.

**TP2-P2.16 MULTIMEDIA INDEXING AND FAST RETRIEVAL BASED ON A VOTE SYSTEM**

*Sylvie Philipp-Foliquet, Guillaume Logerot, ETIS, CNRS, France; Patrick Constant, Pertimm, France; Philippe-Henri Gosselin, ETIS, CNRS, France; Christian Lahanier, C2RMF, France*

We present a new system, called Retimm, for searching databases made of documents containing images and text. Images are indexed by colour and texture distributions. Colour and texture classes are obtained by a quantization adapted to the whole database. Signatures are ranked  $m$  times, once for each dimension, but values are not stored. The search engine works as a vote system : the score for each document is the total of the votes of all coordinates, these last votes depending on a  $k$ -nn search on each dimension. Retimm is able to retrieve very quickly images from large databases from any request composed of one or several images and/or one or several words. The system is interactive, since the query can be modified at any moment by adding or removing images or words.

**TP2-P2.17 LEARNING-BASED INTERACTIVE VIDEO RETRIEVAL SYSTEM**

*Chi-Jiunn Wu, Hui-Chi Zeng, Szu-Hao Huang, Shang-Hong Lai, National Tsing Hua University, Taiwan; Wen-Hao Wang, Industrial Technology Research Institute, Taiwan*

This paper presents an interactive video event retrieval system based on improved adaboost learning. This system consists of three main steps. Firstly, a long video sequence is partitioned into several video clips by using a distribution-based approach instead of detecting shot transition boundaries. Secondly, audiovisual features (i.e., color, motion and audio features) are extracted from video sequences. Finally, the modified AdaBoost learning algorithm is employed for interactive video retrieval with relevance feedback. This AdaBoost learning algorithm differs from conventional AdaBoost learning methods mainly in the selection of paired video features for the weak classifiers. Experimental results show improved performance of video retrieval by using the proposed system.

**TP2-P2.18 SCALABLE IMAGE RETRIEVAL FROM DISTRIBUTED IMAGES DATABASE**

*Tammam Tillo, Marco Grangetto, Gabriella Olmo, Politecnico di Torino, Italy*

In order to store, and retrieve images from large databases, we propose a framework, based on multiple description coding paradigm, that disseminates images over distributed servers. Consequently, decentralized download can be performed, thus reducing links overload and hotspot areas without penalizing downloads speed. Moreover, the tradeoff between system reliability and storage requirement can be achieved by tuning descriptions redundancy, thus providing high flexibility in terms of storage resources, reliability of access, and performance. The scalability of the proposed framework is achieved by the intrinsic progressivity of the multiple description schemes. Moreover, we demonstrate that system can work properly regardless of server crashes.

**TP2-P2.19 A MEASURE FOR EVALUATING RETRIEVAL TECHNIQUES BASED ON PARTIALLY ORDERED GROUND TRUTH LISTS**

*Rainer Typke, Remco C. Veltkamp, Frans Wiering, Universiteit Utrecht, Netherlands*

For a large collection of musical incipits, we have established a ground truth based on the opinions of human experts. It contains correctly ranked matches for a set of given queries. These ranked lists contain groups of documents whose ranks were not significantly different. We introduce the “average dynamic recall” (ADR) measure that averages the recall among a dynamic set of relevant documents. Dynamic recall measures how many of the documents that should have appeared before or at a given position in the result list actually have appeared. ADR at a given position averages this measure up to the given position. Our measure was first used at MIREX 2005.

## **TP2-P2.20 IMAGE REPLICA DETECTION USING R-TREES AND LINEAR DISCRIMINANT ANALYSIS**

*Spyros Nikolopoulos, Stefanos Zafeiriou, Panagiotis Sidiropoulos, Nikos Nikolaidis, Ioannis Pitas, Aristotle University of Thessaloniki, Greece*

In this paper a novel system for image replica detection is presented. The system uses color-based descriptors in order to extract robust features for image representation. These features are used for indexing the images in a database using an R-Tree. When a query about whether a test image is a replica of an image in the database is submitted, the R-Tree is traversed and a set of candidate images is retrieved. Then, in order to obtain a single result and the same time reduce the number of decision errors the system is enhanced with Linear Discriminant Analysis (LDA). The conducted experiments show that the proposed approach is very promising.

## **TP2-P3: Tuesday, July 11, 14:50 - 15:50**

### **TP2-P3 Segmentation (Poster)**

Time: Tuesday, July 11, 14:50 - 15:50

Place: Toronto III

Chair: Alessandro Vinciarelli, IDIAP Research Institute

### **TP2-P3.21 SOCIOMETRY BASED MULTIPARTY AUDIO RECORDINGS SEGMENTATION**

*Alessandro Vinciarelli, IDIAP Research Institute, Switzerland*

This paper shows how Social Network Analysis, the sociological domain studying the interaction between people in specific social environments, can be used to assign roles to different speakers in multiparty recordings. The experiments presented in this work focus on radio news recordings involving around 11 speakers on average. Each of them is assigned automatically a role (e.g. anchorman or guest) without using any information related to their identity or the amount of time they talk. The results (obtained over 96 recordings for a total of around 19 hours) show that more than 85% of the recording time is correctly labeled in terms of role.

### **TP2-P3.22 DETECTION OF VARIANT WIPE EFFECTS**

*Shan Li, Moon-Chuen Lee, The Chinese University of Hong Kong, Hong Kong SAR of China*

Due to the diversity of different wipe effects, wipe transition is considered complex and difficult to detect. This paper identifies two common characteristics of different wipes, which can be described by two principles – independence and completeness. By exploiting the two principles, we developed an effective wipe detector. Since no pre-determined values would be appropriate for different videos, we propose using a dynamic method to generate adaptive thresholds for wipe detection. The experimental results show that the proposed detection method can identify various wipe effects with good accuracy; overall, it outperformed other published methods.

### **TP2-P3.23 EXPERIENTIAL SAMPLING BASED FOREGROUND/BACKGROUND SEGMENTATION FOR VIDEO SURVEILLANCE**

*Pradeep Kumar Atrey, National University of Singapore, Singapore; Vinay Kumar, Anurag Kumar, Indian Institute of Technology, India; Mohan S. Kankanhalli, National University of Singapore, Singapore*

Segmentation of foreground and background has been an important research problem arising out of many applications including video surveillance. A method commonly used for segmentation is “background subtraction” or thresholding the difference between the estimated background image and current image. Adaptive Gaussian mixture based background modelling has been proposed by many researchers for increasing the robustness against environmental changes. However, all these methods, being computationally intensive, need to be optimized for efficient and real-time performance especially at a higher image resolution. In this paper, we propose an improved foreground/background segmentation method which uses Experiential Sampling technique to restrict the computational efforts in the region of interest. We exploit the fact that the region of interest in general is present only in a small part of the image, therefore, the attention should only be focused in those regions. The proposed method shows a significant gain in processing speed at the expense of minor loss in accuracy. We provide experimental results and detailed analysis to show the utility of our method.

**TP2-P3.24 EXTRACTION OF BASEBALL TRAJECTORY AND PHYSICS-BASED VALIDATION FOR SINGLE-VIEW BASEBALL VIDEO SEQUENCES**

*Wei-Ta Chu, Chia-Wei Wang, Ja-Ling Wu, National Taiwan University, Taiwan*

To enrich the viewing experience of baseball games and provide some clues for enhancing pitcher's performance, we propose a Kalman filter-based approach to track ball trajectory from single-view pitching sequences. Without setting extraordinary equipments in stadiums or other sensing instruments, this approach robustly extracts ball trajectory for pitching sequences captured from TV channels or downloaded from the Internet. To validate the detected ball trajectories, we investigate the characteristics of ball trajectories on the basis of a baseball physical model. The effectiveness of ball trajectory extraction and ball position detection are presented.

**TP2-P3.25 AUDIOVISUAL ANCHORPERSON DETECTION FOR TOPIC-ORIENTED NAVIGATION IN BROADCAST NEWS**

*Martin Haller, Technical University of Berlin, Germany; Hyoung-Gook Kim, Samsung Advanced Institute of Technology, Republic of Korea; Thomas Sikora, Technical University of Berlin, Germany*

This paper presents a content-based audiovisual video analysis technique for anchorperson detection in broadcast news. For topic-oriented navigation in newscasts, a segmentation of the topic boundaries is needed. As the anchorperson gives a strong indication for such boundaries, the presented technique automatically determines that high-level information for video indexing from MPEG-2 videos and stores the results in an MPEG-7 conform format. The multimodal analysis process is carried out separately in the auditory and visual modality, and the decision fusion forms the final anchorperson segments.

**TP2-P3.26 VIDEO BREAK DETECTION BASED ON SIMILARITY ANALYSIS AND TEMPORAL STATISTICAL CHARACTERISTICS**

*Jianjun Huang, Fuzhou University, China*

Video Shot boundary detection is a fundamental task in any kind of video content manipulation and retrieval. The area of shot boundary detection has been extensively studied, but achieving highly accurate detection results remains a challenge. In this paper, we present a novel algorithm for video cut detection. The algorithm is implemented and evaluated on the TRECVID benchmark platform. The experimental results show the effectiveness of the proposed approach. The algorithm provides a clue to the gradual transition detection.

**TP2-P3.27 SEMANTIC SEGMENTATION OF DOCUMENTARY VIDEO USING MUSIC BREAKS**

*Aijuan Dong, Honglin Li, North Dakota State University, United States*

Many documentary videos use background music to help structure the content and communicate the semantic. In this paper, we investigate semantic segmentation of documentary video using music breaks. We first define video semantic units based on the speech text that a video/audio contains, and then propose a three-step procedure for semantic video segmentation using music breaks. Since the music breaks of a documentary video are of different semantic levels, we also study how different speech/music segment lengths correlate with the semantic level of a music break. Our experimental results show that music breaks can effectively segment a continuous documentary video stream into semantic units with an average F-score of 0.91 and the lengths of combined segments (speech segment plus the music segment that follows) strongly correlate with the semantic levels of music breaks.

**TP2-P3.28 A NOVEL MODEL-BASED SEGMENTATION APPROACH TO EXTRACT CAPTION CONTENTS ON SPORTS VIDEOS**

*Yih-Ming Su, Chaur-Heh Hsieh, I-Shou University, Taiwan*

The study proposes a novel scheme to extract and recognize the caption contents of various sports captions. A caption extraction process based on an iteratively temporal averaging technique is used to detect and locate a caption region in a series of video frames. Moreover, a caption-content extraction process based on caption identification and model-based segmentation processes is used to extract accurately the contents of various sports captions. Finally, some low-quality character images extracted from the caption contents are recognized using a commercial OCR. Experimental results show that the proposed model-based segmentation approach is very efficient to extract the contents of the various sports captions. Furthermore, the recognition performance from the application of the segmentation approach can be improved about 7.72% in test numeral set, compared to the projection-based segmentation method.



## **TP2-P3.29 FACE SEGMENTATION UNDER UNCONSTRAINED SCENES**

*Jing-Wein Wang, National Kaohsiung University of Applied Sciences, Taiwan*

In this paper, an efficient approach by combining the novel wavelet-based feature template, the support vector machine (SVM) classifier, and the wavelet entropy filtering is presented to robustly detect and segment human face image under complex background. Moreover, a face detection measure (FDM) criterion based on the distance between the expected and the detected eye-mouth triangle circumscribed circle areas is introduced to validate the performance of precise face segmentation.

## **TP2-P3.30 SVM-BASED SHOT BOUNDARY DETECTION WITH A NOVEL FEATURE**

*Kazunori Matsumoto, Masaki Naito, Keiichiro Hoashi, Fumiaki Sugaya, KDDI R&D Laboratories, Inc, Japan*

This paper describes our new algorithm for shot boundary detection and its evaluation. We adopt a 2-stage data fusion approach with SVM technique to decide whether a boundary exists or not within a given video sequence. This approach is useful to avoid huge feature space problems, even when we adopt many promising features extracted from a video sequence. We also introduce a novel feature to improve detection. The feature consists of two kinds of values extracted from a local frame sequence. One is the image difference between the target frame and that synthesized from the neighbors. The other is the difference between neighbors. A fast extraction method for this feature is proposed. Evaluation of our algorithm is conducted with the TRECVID evaluation framework. Our system obtained a high performance at a shot boundary detection task in TRECVID2005.

## **WA1-L1: Wednesday, July 12, 09:50 - 11:10**

### **WA1-L1 Quality of Service Support for Multimedia Transport (Lecture)**

Time: Wednesday, July 12, 09:50 - 11:10

Place: Carmichael

Chair: Yihong Gong, NEC Laboratories America

09:50 - 10:10

### **WA1-L1.1 MEDIA STREAMING WITH CONSERVATIVE DELAY ON VARIABLE RATE CHANNELS**

*Dan Jurca, Pascal Frossard, Ecole Polytechnique Fédérale de Lausanne (EPFL), Switzerland*

We address the problem of delay-constrained streaming of multimedia packets over dynamic bandwidth channels. Efficient streaming solutions generally rely on the knowledge of the channel bandwidth, in order to select the media packets to be transmitted, according with their sending time. However, the streaming server usually cannot have a perfect knowledge of the channel bandwidth, and important packets may be lost because of over-estimation. We address the rate prediction mismatch by media scheduling with a conservative delay, which provides a safety margin for the packet delivery, even in the presence of unpredicted bandwidth variations. We formulate an optimization problem whose goal is to find the optimal conservative delay to be used in the scheduling process, given the network model and the playback delay imposed by the client. We then propose a simple solution to the scheduling delay estimation, effective in real-time streaming scenarios. Our streaming method proves robust against channel prediction errors, and performs better than other mechanisms based on frame reordering strategies.



10:10 - 10:30

**WA1-L1.2 A NOVEL REVERSE FRAME SELECTION SCHEME FOR VIDEO STREAMING OVER VBR CHANNELS**

*Dayong Tao, Jianfei Cai, Nanyang Technological University, Singapore*

In this paper, we propose a Reverse Frame Selection (RFS) scheme based on dynamic programming to solve for the problem of video streaming over VBR channels. In particular, we first consider forward frame selection (FFS) for video streaming over CBR channels. We propose to discard non-optimal states in FFS to reduce the computational cost of dynamic programming. Then we find that the problem can also be solved by RFS with one additional benefit of finding all the optimal results for different preloads in one round. Furthermore, we extend RFS for video streaming over VBR channels where we do not know when and how the channel is going to change in the future. The major advantage of our proposed scheme is that we only need to run RFS several times, and the obtained results can be applied to any type of VBR channels with bandwidth changes occurring at any time. Preliminary results show the good performance of our proposed scheme.

10:30 - 10:50

**WA1-L1.3 SEAMLESS SWITCHING IN MULTI-RATE VIDEO STREAMING SYSTEMS: A WAVELET-BASED SCHEME VERSUS THE SP-FRAME SCHEME**

*Wei Zhang, Bing Zeng, Hong Kong University of Science and Technology, Hong Kong SAR of China*

The multiple bit-rate (MBR) representation of video sequences offers an effective solution to video streaming services over the Internet. To facilitate such MBR-based streaming services, a mechanism is required to support seamless switching among multiple bit-streams when a bandwidth change is detected. The SP-frames developed in H.264 provide such a mechanism at each pre-selected switching point. In this paper, we propose a new switching scheme that is based on a wavelet-domain processing of the reconstructed frame at each switching point. We will compare our scheme with the SP-frame scheme in aspects such as the quality drop at each switching point and all subsequent frames, the count of overhead bits to support an arbitrary switching, and the computational complexity. The results indicate that our scheme can achieve the seamless switching with a better rate-distortion performance at cost of slightly increasing the computations at the decoder side.

**WA1-L2: Wednesday, July 12, 09:50 - 11:10**

**WA1-L2 Media Processing and Algorithms (Lecture)**

Time: Wednesday, July 12, 09:50 - 11:10

Place: Johnston

Chair: Feng Wu, Microsoft Research Asia

09:50 - 10:10

**WA1-L2.1 COMBINING TEXTUAL AND VISUAL ONTOLOGIES TO SOLVE MEDICAL MULTIMODAL QUERIES**

*Saïd Radhouani, University of Geneva, Switzerland; Joo-Hwee Lim, Jean-Pierre Chevallet, Institute for Infocomm Research, Singapore; Gilles Falquet, University of Geneva, Singapore*

In order to solve medical multimodal queries, we propose to split the queries in different dimensions using ontology. We extract both textual and visual terms depending on the ontology dimension they belong to. Based on these terms, we build different sub queries each corresponds to one query dimension. Then we use Boolean expressions on these sub queries to filter the entire document collection. The filtered document set is ranked using the techniques in Vector Space Model. We also combine the ranked lists generated using both text and image indexes to further improve the retrieval performance. We have achieved the best overall performance for the Medical Image Retrieval Task in CLEF 2005. These experimental results show that while most queries are better handled by the text query processing as most semantic information are contained in the medical text cases, both textual and visual ontology dimensions are complementary in improving the results during media fusion.

10:10 - 10:30

**WA1-L2.2 EFFICIENT HAND GESTURE RENDERING AND DECODING USING A SIMPLE GESTURE LIBRARY**

*Jason Smith, Lijun Yin, State University of New York at Binghamton, United States*

Recent work in hand gesture rendering and decoding has treated the two fields as separate and distinct. As the work of rendering evolves, it emphasizes exact movement replication, including more muscle and skeletal parameterization. The work in gesture decoding is largely centered on trained systems, which require large amounts of time in front of a camera rendering a gesture in order to decode movement. This paper presents a new scheme which more tightly couples the gesture rendering and decoding processes. While this scheme is simpler than existing techniques, the rendering remains natural looking, and decoding a new gesture does not require extensive training.

10:30 - 10:50

**WA1-L2.3 MOTION ESTIMATION BY QUADTREE PRUNING AND MERGING**

*Marco Tagliasacchi, Mauro Sarchi, Stefano Tubaro, Politecnico di Milano, Italy*

In this paper we propose a rate-distortion optimized motion estimation algorithm that is built upon a quadtree structure. Each node of the quadtree represents a block in the current frame together with its motion vector, and the block size decreases from the root to the leaves. In the first step, the quadtree is pruned according to a rate-distortion criterion in order to obtain blocks of variable sizes. A further rate rebate can be achieved by merging those leaf nodes of the quadtree that can be efficiently represented by the same motion vector. The proposed merging scheme provides a reduction of up to 50% of the rate spent for the motion model with respect to the case that performs pruning only.

10:50 - 11:10

**WA1-L2.4 ADAPTIVE MCTF BASED ON CORRELATION NOISE MODEL FOR SNR SCALABLE VIDEO CODING**

*Ruiqin Xiong, Chinese Academy of Sciences, China; Jizheng Xu, Feng Wu, Shipeng Li, Microsoft Research Asia, China*

This paper proposes a subband adaptive motion compensated temporal filtering (MCTF) technique for scalable video coding and introduces a revised synthesis gain model for the quantization in this adaptive MCTF scheme. In scalable video coding, hierarchical MCTF is extensively adopted to exploit the temporal correlation across video frames. In this hierarchical MCTF structure, the strength of temporal correlation varies with the level of temporal transform and varies with the various spatial frequency components in a frame. The reconstruction noises also have diverse strength at various subbands. According to the correlation and noise characteristics of various subbands, we can adjust the strength of motion compensated prediction step in MCTF to maximally take the advantage of temporal correlation but restrict the propagation of reconstruction noise. The quantization step of each subband is also adjusted according to synthesis gain determined by the MCTF structure. In this way an adaptive MCTF scheme is formed and the proposed technique improves the coding performance of scalable video coding.

**WA1-L3 3D-TV: Primed for Success? II (Special Session)**

Time: Wednesday, July 12, 09:50 - 11:10

Place: Thomson

Co-Chairs: Liang Zhang, Communications Research Centre Canada, Wa James Tam, Communications Research Centre Canada and Hirokazu Yamanoue, NHK Science &amp; Technical Research Laboratories

09:50 - 10:10

**WA1-L3.1 3D-TV CONTENT GENERATION: 2D-TO-3D CONVERSION***Wa James Tam, Liang Zhang, Communications Research Centre Canada, Canada*

The next major advancement in television is expected to be stereoscopic three-dimensional television (3D-TV). A successful roll-out of 3D-TV will require a backward-compatible transmission and distribution system, inexpensive 3D displays that are equal or superior to high-definition television (HDTV), and an adequate supply of high-quality 3D program material. With respect to the last factor, we reckon that the conversion of 2D material to stereoscopic 3D could play an important role. In this paper we provide (a) an overview of the fundamental principle underlying 2D-to-3D conversion techniques, (b) a cursory look at a number of approaches for depth extraction using a single image, and (c) a highlight of the potential use of surrogate depth maps in depth image based rendering for 2D-to-3D conversion. This latter approach exploits the ability of the human visual system to combine reduced disparity information that are located mainly at edges and object boundaries with pictorial depth cues to produce an enhanced sensation of depth over 2D images.

10:10 - 10:30

**WA1-L3.2 DIBR-BASED 3D VIDEOS USING NON VIDEO RATE RANGE IMAGE STREAM***Xiaoyi Jiang, University of Münster, Germany; Martin Lambers, University of Siegen, Germany*

The fundamental assumption of 3D videos using depth-image-based rendering is the full availability of range images at video rate. In this work we alleviate this hard demand and assume that only limited resources of range images are available, i.e. corresponding range images exist for some, but not all, color images of the monoscopic video stream. We propose to synthesize the missing range images between two consecutive range images. Experiments on real videos have demonstrated very encouraging results. Especially, one 3D video was generated from a 2D video without any sensory 3D data available at all. In a quality evaluation using an autostereoscopic 3D display the test viewers have attested similar 3D video quality for our synthesis technique and rendering based on depth ground truth.

10:30 - 10:50

**WA1-L3.3 A FLEXIBLE 3D TV SYSTEM FOR DIFFERENT MULTI-BASELINE GEOMETRIES***Oliver Schreer, Christoph Fehn, Nicole Atzpadin, Marcus Müller, Aljoscha Smolic, Ralf Tanger, Peter Kauff, Fraunhofer – Heinrich Hertz Institute, Germany*

Interoperability, scalability and adaptability are important features for a successful introduction of future 3D TV services. Hence, new concepts must be able to adapt the multi-view geometry of the capturing system to the geometry of the 3D reproduction systems. An approach is discussed, which considers these adaptation issues based on the concept of an  $N \times$  video-plus-depth data representation. The core algorithms for depth map creation on the analysis side and depth image based rendering on the reproduction side are presented.

10:50 - 11:10

**WA1-L3.4 NEAR-FUTURE STREAMING FRAMEWORK FOR 3D-TV APPLICATIONS***Goran Petrovic, Peter H.N. de With, University of Technology Eindhoven, Netherlands*

This paper presents a layered framework for 3D-TV applications, combining multiview and depth-image based approaches in a scalable fashion. To solve the problem of missing data due to disocclusions, we add specific layers for coded occlusion data and the edge-mask information for high-quality 3D rendering of key objects in the scene. We show how the same framework can be extended towards FTV applications by jointly addressing simulcast and multicast transmission. By adopting a distributed delivery architecture, new interesting properties can be realized such as shared processing for the creation and streaming of virtual viewpoints.

**WA1-L4 Applications to Sports Video (Lecture)**

Time: Wednesday, July 12, 09:50 - 11:10

Place: Toronto I

Chair: Changsheng Xu, Institute for Infocomm Research

09:50 - 10:10

**WA1-L4.1 SHOT AND SCORING EVENTS IDENTIFICATION OF BASKETBALL VIDEOS***Chung-Lin Huang, Huang-Chia Shih, Ching-Lun Chen, National Tsing Hua University, Taiwan*

This paper presents a novel semantic-oriented video analysis system for the basketball game videos. Based on Bayesian Belief Network (BBN), it may bridge this gap between the low-level features describing image/video structure and the high-level knowledge. We apply the Support Vector Machine (SVM) to identify and track the ball, the shooter, and the basket as the low-level features. Based on these features, our BBN framework can identify four categories of shot event such as short shot, medium shot, long shot, free throw, and the scoring event. In the experiments, we demonstrate that our system may interpret the video shots in terms of four different shot events and one scoring event effectively.

10:10 - 10:30

**WA1-L4.2 A SYSTEM FOR AUTOMATIC JUDGMENT OF OFFSIDES IN SOCCER GAMES***Sadatsugu Hashimoto, Shinji Ozawa, Keio University, Japan*

In this paper, we propose a system for automatic judgment of offsides in soccer games. We detect and track players in fixed multi camera images and calculate the world coordinates of them. Furthermore, we do a formation analysis by classifying uniforms and calculate the position of an offside line. On the other hand, we calculate the 3D coordinates and the trajectories of a ball in world coordinates from the plane coordinates of a ball in multi cameras and recognize the moment of a play from the 3D trajectories of a ball. In addition, we make a judge player's interfering with play by analyzing the spatial relationship between a ball and players. Finally, we make an offside judgment by integrating these results. We apply our system to a real soccer match and demonstrate the availability of this system by showing the experimental results.

10:30 - 10:50

**WA1-L4.3 AUTOMATIC CONTENT PLACEMENT IN SPORTS HIGHLIGHTS***Kongwah Wan, Changsheng Xu, Institute for Infocomm Research, Singapore*

To be viable advertising platforms, methods for in-program content placement in sports video must balance against clutter. We propose viewer relevance (VR) measures of video frames in the temporal and spatial domain. Video sub-segments with low temporal VR are first selected, within which actual content is emplaced on regions with low spatial VR. We compute VR measures using color, motion, texture and domain features, upon which spatio-temporal techniques are used to segment spatial regions for actual content placement. Results from preliminary subjective viewing trials on soccer and tennis video indicate that our approach is promising.

10:50 - 11:10

**WA1-L4.4 FULLY AND SEMI-AUTOMATIC MUSIC SPORTS VIDEO COMPOSITION***Jinjun Wang, Engsiong Chng, Nanyang Technological University, Singapore; Changsheng Xu, Institute for Infocomm Research, Singapore*

Video composition is important for music video production. In this paper we propose an automatic method to assist the music sports video composition operation. Our approach is based on Dynamic Programming algorithm which finds a set of video shots that best matches the music. The method by default is fully-automatic, and users specification could be inserted to control the composition process, making it a semiautomatic system. This research has obvious importance to reduce manual processing, and enables the generation of high quality personalized music sports video. The proposed method is generic and fast. The experimental results are satisfactory.

**WA1-L5 Feature Extraction and Segmentation III (Lecture)**

Time: Wednesday, July 12, 09:50 - 11:10

Place: Toronto II

Chair: Thomas Huang, University of Illinois, Urbana-Champaign

09:50 - 10:10

**WA1-L5.1 IMPROVING SPEAKER DIARIZATION BY CROSS EM REFINEMENT**

*Huazhong Ning, University of Illinois at Urbana-Champaign, United States; Wei Xu, Yihong Gong, NEC Laboratories America, Inc., United States; Thomas S. Huang, University of Illinois at Urbana-Champaign, United States*

In this paper, we present a new speaker diarization system that improves the accuracy of traditional hierarchical clustering-based methods with little increase in computational cost. Our contributions are mainly two fold. First, we include a preprocessing called “local clustering” before the hierarchical clustering algorithm to merge very similar adjacent speech segments. This local clustering aims to reduce the number of segments to be clustered by the hierarchical clustering, so as to dramatically increase the processing speed. Second, we perform a postprocessing called “cross EM refinement” to purify the clusters generated by the hierarchical clustering. This algorithm is based on the idea of cross validation and EM algorithm. Our experimental evaluations show that the proposed cross EM refinement approach reduces the speaker diarization error by up to 56%, with an average reduction of 22% compared to the traditional hierarchical clustering method.

10:10 - 10:30

**WA1-L5.2 IMAGE SEGMENTATION USING PARALLEL SELF ORGANIZING TREE MAP**

*Xiaoming Fan, Ryerson University, Canada; Jonathan Randall, University of Sydney, Australia; Ivan Lee, Ryerson University, Canada*

At present, the researchers are making great effort to search for a general purpose resolution for image segmentation. Most of the approaches we are using now are somewhat inefficient and tardy. In this paper we present an innovative method, Parallel Self-organizing Tree Map (PSOTM), which is improved from SOTM. By processing image in parallel, PSOTM could segment the image in a much faster processing speed. After applying the new technology, PSOTM, we could obtain the higher efficiency with limited impacts on the visual quality.

10:30 - 10:50

**WA1-L5.3 MOTION SEGMENTATION OF 3D VIDEO USING MODIFIED SHAPE DISTRIBUTION**

*Toshihiko Yamasaki, Kiyoharu Aizawa, University of Tokyo, Japan*

In this paper, temporal segmentation of 3D video based on motion analysis is presented. 3D video is a sequence of 3D models made for a real-world dynamic object. A modified shape distribution algorithm is proposed to realize stable shape feature representation. In our approach, representative points are generated by clustering vertices based on their spatial distribution instead of randomly sampling vertices as in the original shape distribution algorithm. Motion segmentation is conducted analyzing local minima in degree of motion calculated in the feature vector space. The segmentation algorithm developed in this paper does not require any predefined threshold values but rely on relative relationships among local minima and local maxima of the motion. Therefore, robust segmentation has been achieved. The experiments using 3D video of traditional dances yielded encouraging results with the precision and recall rates of 93% and 88%, respectively, on average.

10:50 - 11:10

**WA1-L5.4 FEMA: A FAST EXPECTATION MAXIMIZATION ALGORITHM BASED ON GRID AND PCA**

*Zhiwen Yu, Hau-San Wong, City University of Hong Kong, Hong Kong SAR of China*

EM algorithm is an important unsupervised clustering algorithm, but the algorithm has several limitations. In this paper, we propose a fast EM algorithm (FEMA) to address the limitations of EM and enhance its efficiency. FEMA achieves low running time by combining principal component analysis(PCA), a grid cell expansion algorithm(GCEA) and a hierarchical cluster tree. PCA and multi-dimensional grid are applied to find a set of “good” initial parameters for the EM algorithm, while the hierarchical cluster tree deals with the case where the cluster is concave by making use of a merging algorithm. The experiments indicate that FEMA outperforms EM by reducing 45% of the CPU time.

**WA1-L6 Coding Algorithms and Standards III (Lecture)**

Time: Wednesday, July 12, 09:50 - 11:10

Place: Varley

Chair: Xiaolin Wu, McMaster University

09:50 - 10:10

**WA1-L6.1 A COMBINED BAYESSHRINK WAVELET-RIDGELET TECHNIQUE FOR IMAGE DENOISING***Nezamoddin N. Kachouie, Paul Fieguth, University of Waterloo, Canada*

In this paper a combined BayesShrink Wavelet-Ridgelet denoising method is presented. In our previous work we have showed that BayesShrink Ridgelet performs better than VisuShrink Ridgelet and VisuShrink Wavelet. Although our BayesShrink Ridgelet technique performs somewhat poorer in comparison with BayesShrink Wavelet, based on SNR, visually it produces smoother results, especially for images with straight lines. In the proposed method BayesShrink Wavelet is combined with BayesShrink Ridgelet denoising method which performs better than each filter individually. The proposed combined denoising method gains the advantage of each filter in its specific domain, i.e., Wavelet for natural and Ridgelet for straight regions, and produces better and smoother results, both visually and in terms of SNR.

10:10 - 10:30

**WA1-L6.2 ERROR RESILIENT MULTIPLE DESCRIPTION COMPRESSION OF VECTOR GRAPHICS***Martin Röder, University of Konstanz, Germany; Xiaolin Wu, Sorina Dumitrescu, McMaster University, Canada*

This research is motivated by the needs of robust streaming of vector graphics contents over the Internet, wireless and other lossy networks. We present a multiple description coding (MDC) technique for error resilient compression and transmission of 2D vector graphics contents. An object is coded into two or more so-called co-descriptors, which are transmitted in separate data packets and generally via different network routes from a server to a client. Each co-descriptor can autonomously provide an approximation of the input object, and it can collaborate with other co-descriptors, if also available at the decoder, to refine the approximation.

10:30 - 10:50

**WA1-L6.3 MULTIPLIERLESS APPROXIMATION OF FAST DCT ALGORITHMS***Raymond K.W. Chan, Moon-Chuen Lee, The Chinese University of Hong Kong, Hong Kong SAR of China*

This paper proposes an effective method for converting any fast DCT algorithm into an approximate multiplierless version. Basically it approximates any constant in the original transform by a signed digit representation. We developed an efficient algorithm to convert any constant into a signed digit string with a minimum number of non-zero signed digits and a reduced length. As the accuracy of an approximated algorithm depends critically on the assignment of signed digits to the constants, this paper formulated an effective algorithm for finding an effective signed digits configuration which could minimize the MSE of an approximated DCT algorithm with a specified complexity. Experiment results show that the AAN's fast DCT algorithm, approximated by the proposed method and using an optimized configuration can be used to reconstruct images with high visual quality in terms of PSNR.

10:50 - 11:10

**WA1-L6.4 ANALYSIS OF HIERARCHICAL B PICTURES AND MCTF***Heiko Schwarz, Detlev Marpe, Thomas Wiegand, Fraunhofer Institute for Telecommunications - Heinrich Hertz Institute, Germany*

In this paper, an investigation of H.264/MPEG4-AVC conforming coding with hierarchical B pictures is presented. We analyze the coding delay and memory requirements, describe details of an improved encoder control, and compare the coding efficiency for different coding delays. Additionally, the coding efficiency of hierarchical B picture coding is compared to that of MCTF-based coding by using identical coding structures and a similar degree of encoder optimization. Our simulation results turned out that in comparison to the widely used IBBP... structure coding gains of more than 1 dB can be achieved at the expense of an increased coding delay. Further experiments have shown that the coding efficiency gains obtained by using the additional update steps in MCTF coding are generally smaller than the losses resulting from the required open-loop encoder control.

**WA2-L1 Application Driven Multimedia Systems and Architectures (Lecture)**

Time: Wednesday, July 12, 11:10 - 12:30

Place: Carmichael

Chair: Liang-Gee Chen, National Taiwan University

11:10 - 11:30

**WA2-L1.1 ADAPTATION SCENARIOS FOR NEW MEDIA ARTWORKS**

*Anis Ouali, Brigitte Kerhervé, Odile Marcotte, Paul Landon, Université du Québec À Montréal, Canada*

Artists in new media arts explore the possibilities offered by digital technologies to create adaptive and interactive new media artworks integrating audio and video. In this paper, we are interested in adaptation mechanisms for such artworks and we focus on the design and implementation of adaptation scenarios. We formalize adaptation scenarios by using the event-condition-action paradigm and we introduce the concept of adaptation policy. We present the adaptation engine we have designed and prototyped to manage and execute adaptation scenarios.

11:30 - 11:50

**WA2-L1.2 FETAL HEART SIGNAL MONITORING WITH CONFIDENCE FACTOR**

*Jianfeng Chen, Koksoon Phua, Ying Song, Louis Shue, Institute for Infocomm Research, Singapore*

In this paper we present a portable PDA-based fetal heart monitor that is able to provide instantaneous fetal heart rate (FHR) for the pregnant women. Several signal processing techniques are implemented in our device to make it clear, reliable and ease-to-use. Firstly the spectral subtraction method is used to remove the stationary background noise and the amplitude modulation technique is then employed to make the fetal heart sound more audible. A real time algorithm is proposed to extract the FHR information from the weak and noisy signals. Based on the signal spectrum distribution, a confidence factor (CF) is designed to forecast the reliability of the calculated FHR. CF can also be used to guide the user in locating the optimum auscultation position. Our device has been tested in a local hospital on a sample group of 41 pregnant women. It has been shown that an approximate 75% of the pregnant mothers are able to hear their babies' heart sound clearly by using our device. Tests are also carried out to compare our resulted FHR with the cardiotocography (CTG) records.

11:50 - 12:10

**WA2-L1.3 LOW POWER ENTROPY CODING HARDWARE DESIGN FOR H.264/AVC BASELINE PROFILE ENCODER**

*Chuan-Yung Tsai, Tung-Chien Chen, Liang-Gee Chen, National Taiwan University, Taiwan*

Low power hardware design for entropy coding of H.264/AVC baseline profile encoder is urgent for the increasing mobile applications. However, previous works are poor in the power performance. In this paper, the first low power Context-based Adaptive Variable Length Coding (CAVLC) scheme named the Side Information Aided (SIA) Symbol Look Ahead (SLA) one-pass CAVLC is proposed, with the non-zero and abs-one SIA flags. A reconfigurable architecture for the SLA module is also proposed to support the low power CAVLC scheme efficiently. The resultant hardware power is reduced by 69% to only 3.7mW at 27MHz and 1.8V for CIF-sized video coding. The total logic gate count is 27K gates.

12:10 - 12:30

**WA2-L1.4 USER-ASSISTED IMAGE CLASSIFICATION ON PERSONAL PHOTO COLLECTIONS**

*Wei-Ta Chen, Ying-Hsiang Wen, Ming-Syan Chen, National Taiwan University, Taiwan*

Image classification on personal photo collections can be extremely useful to various management tasks. However, there is little progress made forwarding it due to (1) lack of training data, and (2) subjectivity inherent in a user's photo-organizing behavior. In this paper, we propose a framework, User-assisted image Classification on Personal Photo Collections (UCP), to address this problem. The uniqueness of this framework is that it is user-centric and includes users in the loop. Our experimental results show that the techniques used in this framework are promising.



**WA2-L2     Multimedia Systems Engineering: System Modeling, Algorithm, Simulation, and Analysis (Lecture)**

Time:        Wednesday, July 12, 11:10 - 12:30

Place:       Johnston

Chair:       James Gordy, Carleton University

11:30 - 11:50

**WA2-L2.2     JOINT RATE-DISTORTION-COMPLEXITY OPTIMIZATION FOR H.264 MOTION SEARCH**

*Yu Hu, Qing Li, Siwei Ma, C.-C. Jay Kuo, University of Southern California, United States*

A joint rate-distortion-complexity H.264 motion search framework is proposed to balance the encoder's coding efficiency and complexity in an embedded system environment. Under our framework, the complexity of H.264 motion search is primarily measured by the execution time of the sum of absolute differences (SAD) calculation. Two Lagrange parameters are used to terminate the complexity-inefficient motion search rounds and skip redundant motion search of small block modes, respectively. Then, the relationship between the weighted complexity and the Lagrange parameters is explored to allocate the complexity cost among different coding units. It is demonstrated by experimental results that the proposed method can reduce the complexity without much sacrifice in coding efficiency.

11:50 - 12:10

**WA2-L2.3     POSTFILTERING FOR SUPPRESSION OF RESIDUAL ECHO FROM VOCODER DISTORTION IN PACKET-BASED TELEPHONY**

*James Gordy, Rafik Goubran, Carleton University, Canada*

This paper investigates postfiltering for residual echo suppression in networks employing low-bit-rate speech compression in the echo path. Simulations show that the residual echo from nonlinear vocoder distortion with ITU G.729 is proportional to the input signal LPC spectrum. An algorithm is proposed to estimate the residual echo power spectrum using a frequency-dependent scaling factor. The algorithm is incorporated into a psychoacoustic postfilter for residual echo suppression and compared to an existing estimator with a fixed scaling factor. Experiments with speech input and near-end signals show an average 0.85 dB lower spectral distortion and 0.4 higher estimated mean opinion score.

12:10 - 12:30

**WA2-L2.4     ONLINE DOUBLETALK DETECTOR CALIBRATION FOR ACOUSTIC ECHO CANCELLATION IN VIDEOCONFERENCING SYSTEMS**

*James Gordy, Rafik Goubran, Carleton University, Canada*

This paper addresses the problem of doubletalk detector calibration for acoustic echo cancellers in hands-free environments such as videoconferencing. A statistical model of a recently proposed doubletalk detector is used to show that optimal detection thresholds are dependent on the input signal and the adaptive filter error. A signal-adaptive algorithm is proposed for calculating an optimal threshold for arbitrary input signals and echo path environments. Simulation results verify the improvement in detection probability offered by the proposed algorithm compared to simple empirical calibration methods.

**WA2-L3 Multimedia over Wireless Networks (Lecture)**

Time: Wednesday, July 12, 11:10 - 12:30

Place: Thomson

Chair: Hayder Radha, Michigan State University

11:10 - 11:30

**WA2-L3.1 UNEQUAL ERROR PROTECTION FOR VIDEO STREAMING OVER WIRELESS LANS USING CONTENT-AWARE PACKET RETRY LIMIT***Chih-Ming Chen, National Tsing Hua University, Taiwan; Chia-Wen Lin, National Chung Cheng University, Taiwan; Yung-Chang Chen, National Tsing Hua University, Taiwan*

In this paper, we propose a content-aware retry limit adaptation scheme for video streaming over IEEE 802.11 wireless LANs (WLANs). Video packets of different importance are unequally protected with different retry limits at the MAC layer. The loss impact of each packet is estimated to guide the selection of its retry limit. More retry numbers are allocated to packets of higher loss impact to achieve unequal error protection. Experimental results show that the proposed adaptation scheme can effectively mitigate the error propagation due to packet loss and assure the on-time arrival of packets for presentation, thereby improving video quality significantly.

11:30 - 11:50

**WA2-L3.2 ENERGY-AWARE MULTI-SOURCE VIDEO STREAMING***Danjue Li, Chen-Nee Chuah, University of California, Davis, United States; Gene Cheung, HP Laboratories, United States; S. J. Ben Yoo, University of California, Davis, United States*

In a multi-source video streaming system, premature draining of low-power nodes can cause sudden failures of peer connections and degrade streaming performance. To solve this problem, we propose an energy-aware scheduling (EAS) scheme to better distribute the streaming load among different peers by jointly considering network conditions and node energy levels. We model the proposed scheme using a rate/energy-distortion optimization framework and heuristically solve it using the concept of asynchronous clocks. Simulation studies show that the proposed EAS scheme can achieve comparable streaming quality while consuming less energy.

11:50 - 12:10

**WA2-L3.3 PACKET SCHEDULING OF STREAMING VIDEO WITH FLEXIBLE REFERENCE FRAME USING DYNAMIC PROGRAMMING AND INTEGER ROUNDING***Gene Cheung, Wai-tian Tan, HP Laboratories, Japan*

Video coding standards like H.264 offer the flexibility to select reference frames during motion estimation for predicted frames. We investigate the packet scheduling problem of streaming video over lossy networks from a real-time encoder with flexible reference frame. In particular, we consider a multi-path streaming setting where each predicted frame of video, in addition to the flexibility to select a reference frame, can schedule one or multiple transmissions on one or multiple delivery paths for the upcoming optimization period. We present an algorithm based on dynamic programming that provides a locally optimal solution with high complexity. We then present a rounding method to reduce computation complexity at the expense of degrading solution quality. Results show that our algorithm performs noticeably better than a greedy scheme, and graceful tradeoff between complexity and solution quality can be achieved.

12:10 - 12:30

**WA2-L3.4 UTILIZING SSR INDICATIONS FOR IMPROVED VIDEO COMMUNICATION IN PRESENCE OF 802.11B RESIDUE ERRORS***Shirish Karande, Utpal Parrikar, Kiran Misra, Hayder Radha, Michigan State University, United States*

Radio hardware used for the reception of 802.11b frames is capable of associating a Signal to Silence Ratio (SSR) with each received frame. If a received frame is corrupted, then these SSR indications can be used to provide robust apriori estimate of the bit error rate in the packet. In many recently proposed cross-layer protocols, for transmission of video over wireless networks, recovery of information from partially corrupted packets has shown significant utility. In this paper, based on experiments with actual 802.11b error traces, we show that the Channel State Information (CSI) provided by the SSR indications can be used to improve the error recovery performance of an FEC scheme employed in conjunction with a cross-layer protocol. H.264 based simulation are used to establish the efficacy of the proposed work for video applications; specifically for video over 802.11b WLAN.

**WA2-L4 Music Indexing and Retrieval (Lecture)**

Time: Wednesday, July 12, 11:10 - 12:30

Place: Toronto I

Chair: Suh-Yin Lee, National Chiao Tung University

11:10 - 11:30

**WA2-L4.1 DETECTING CHANGES IN USER-CENTERED MUSIC QUERY STREAMS***Hua-Fu Li, National Chiao-Tung University, Taiwan; Man-Kwan Shan, National Chengchi University, Taiwan; Suh-Yin Lee, National Chiao-Tung University, Taiwan*

In this paper, we propose an efficient algorithm, called MQS-change (changes of Music Query Streams), to detect the changes of maximal melody structures in user-centered music query streams. Two music melody structures (set of chord-sets and string of chord-sets) are maintained and four melody structure changes (positive burst, negative burst, increasing change and decreasing change) are monitored in a new data structure MSC-list (a list of Music Structure Changes). Experiments show that MQS-change algorithm is an online, single-pass approach to detect the changes of music melody structures over continuous music query streams.

11:30 - 11:50

**WA2-L4.2 AN EFFICIENT NUMERIC INDEXING TECHNIQUE FOR MUSIC RETRIEVAL SYSTEM***Chuan-Wang Chang, Hewijin Christine Jiau, National Cheng Kung University, Taiwan*

Space requirement for storing indexes and performance for query processing are two critical issues in music information retrieval (MIR) system. To overcome difficulties in variable length of queries and enhance efficiency of music retrieval, we propose an effective and efficient numeric indexing structure. It differs greatly from pre-existing researches in textual indexing techniques. We show how the development of this framework has been motivated and demonstrate how the technique may be naturally applied to solve this two fundamental MIR issues. Experiments are performed to compare our method with previous solutions. The results show that our method is more scalable and economical than previous methods. The method we proposed can achieve dramatically and significantly improvement in saving time and storing space for retrieving and indexing.

11:50 - 12:10

**WA2-L4.3 ONLINE MINING OF RECENT MUSIC QUERY STREAMS***Hua-Fu Li, Chin-Chuan Ho, National Chiao-Tung University, Taiwan; Man-Kwan Shan, National Chengchi University, Taiwan; Suh-Yin Lee, National Chiao-Tung University, Taiwan*

Mining multimedia data is one of the most important issues in data mining. In this paper, we propose an online one-pass algorithm to mine the set of frequent temporal patterns in online music query streams with a sliding window. An effective bit-sequence representation is used to reduce the processing time and memory needed to slide the windows. Experiments show that the proposed algorithm only needs a half of memory requirement of original music query data, and just scans the data once.

12:10 - 12:30

**WA2-L4.4 A FINGERPRINTING SYSTEM FOR MUSICAL CONTENT***Lahouari Ghouti, Ahmed Bouridane, Queen's University of Belfast, United Kingdom; Mohammad Ibrahim, De Montfort University, United Kingdom*

Digital multimedia content (especially audio) is becoming a major part of the average computer user experience. Large digital audio collections of music, audio and sound effects are also used by the entertainment, music, movie and animation industries. Therefore, the need for identification and management of audio content grows proportionally to the increasing widespread availability of such media virtually "any time and any where" over the Internet. In this paper, we propose a novel framework for musical content fingerprinting using balanced multiwavelets (BMW). The framework for generating robust perceptual fingerprint (or hash) values is described. The generated fingerprints are used for identifying, searching, and retrieving audio content from large digital music databases. Furthermore, we illustrate, through extensive computer simulation, the robustness of the proposed framework to efficiently represent musical content and withstand several signal processing attacks and manipulations.

**WA2-L5 Multimedia for Healthcare (Special Session)**

Time: Wednesday, July 12, 11:10 - 12:30

Place: Toronto II

Co-Chairs: Datong Chen and Jie Yang

11:10 - 11:30

**WA2-L5.1 APPLICATIONS OF AUTOSTEREOSCOPIC DISPLAYS IN OPHTHALMOLOGIC STUDIES***Xiaoyi Jiang, University of Münster, Germany; Daniel Mojon, Kantonsspital St.Gallen, Switzerland*

Autostereoscopic displays are an emerging technology which provide 3D viewing experiences without the need of glasses or other encumbering viewing aids. In this paper we explore their potential in ophthalmologic studies. While the vast majority of applications of autostereoscopic displays in other fields is merely based on producing 3D viewing effects, we can distinguish between several classes of ophthalmologic tasks in which autostereoscopic displays play a very different role. Three concrete applications are described. With the steady improvements in autostereoscopic displays we expect to develop qualitatively new ophthalmologic tests in future.

11:30 - 11:50

**WA2-L5.2 LABEL DISAMBIGUATION AND SEQUENCE MODELING FOR IDENTIFYING HUMAN ACTIVITIES FROM WEARABLE PHYSIOLOGICAL SENSORS***Wei-Hao Lin, Alexander Hauptmann, Carnegie Mellon University, United States*

Wearable physiological sensors can provide a faithful record of a patient's physiological states without constant attention of caregivers. A computer program that can infer human activities from physiological recordings will be an valuable tool for physicians. In this paper we investigate to what extent current machine learning algorithms can correctly identify human activities from physiological sensors. We further identify two challenges that developers need to address. The first problem is that the labels of training data are inevitably noisy due to difficulties of annotating thousands hours of data. The second problem lies in the continuous nature of human activities, which violates the independence assumption made by many learning algorithms. We approach the first problem of noisy labeling in the multiple-label framework, and develop a conditional Markov Models to take temporal context into consideration. We evaluate the proposed methods on 12,000 hours of the physiological recordings. The results show that Support Vector Machines are effective to identify human activities from physiological signals, and efforts of disambiguating noisy labels are worthwhile.

11:50 - 12:10

**WA2-L5.3 SPECIALIZED VIDEO AND PHYSIOLOGICAL DATA CODING SYSTEM FOR REMOTE MONITORING***Qiang Liu, University of Pittsburgh, United States; Datong Chen, Carnegie Mellon University, United States; Mingui Sun, Robert J. Scabassi, University of Pittsburgh, United States*

Patient monitoring via video and physiological data recording can now be performed outside hospitals. This procedure, usually performed in a prolonged manner, generates a considerable amount of data, which calls for efficient ways for archiving and transmission. In this work, we present a specialized system to code the video and the physiological data recorded from a patient, aiming at a reduced bandwidth requirement compared to the conventional methods. We've developed an object-based approach to coding the monitoring video. By applying two change detection methods, we decompose a video frame into three video object planes (VOPs) representing the background, the stationary foreground and the moving foreground. These VOPs are coded at different frame rates, leading to a reduced overall bit rate. For coding the physiological data (using electroencephalogram, i.e. EEG, as an example), we present an effective solution by using a combination of the lifting scheme and the SPIHT algorithm. This approach is featured with a wavelet-quantization algorithm that enables a scalable transmission. The feasibility of this proposed system is demonstrated by our experimental results.

12:10 - 12:30

**WA2-L5.4 AUTOMATIC MEDICAL IMAGE ANNOTATION AND RETRIEVAL USING SEMI-SECC***Jian Yao, Zhongfei Zhang, Binghamton University, United States; Sameer Antani, Rodney Long, George Thoma, National Institutes of Health, United States*

The demand for automatically annotating and retrieving medical images is growing faster than ever. In this paper, we present a novel medical image annotation method based on the proposed Semantic Error-Correcting output Codes (SECC). With this annotation method, we present a new semantic image retrieval method. The experimental results on IMAGECLEF 2005 annotation data set clearly show the strength and the promise of the presented methods.

**WA2-L6 Coding Algorithms and Standards IV (Lecture)**

Time: Wednesday, July 12, 11:10 - 12:30

Place: Varley

Chair: Jacky Shen, Microsoft Research Asia

11:10 - 11:30

**WA2-L6.1 H.263 VIDEO CODEC PERFORMANCE WITH A FAST 8X8 INTEGER IDCT***Joao Tavares, Antonio Silva, Antonio Navarro, Aveiro University, Portugal*

Several standardized video coding algorithms use a well known discrete cosine transform (DCT) at the encoder to remove redundancy from video random processes. Its inverse, the IDCT is present at the decoder as well as at the encoder loop. The accuracy of this inverse transform, required to avoid large drift between the encoder and decoder, is defined in an IEEE standard which will be withdrawn from MPEG and ITU standardized codecs. Due to the huge number of computations required to compute the FDCT/IDCT (Forward/Inverse DCT) pair, reduction of their complexity is essential to speed up the video processing. In this paper, we propose a fast integer IDCT calculation method. Additionally, we insert it into the H.263 reference software in order to validate our proposed method. The testing results using two different video sequences, at QCIF and CIF resolutions, show similar PSNR average values between the reference H.263 and the proposed H.263 codec.

11:30 - 11:50

**WA2-L6.2 SYNDROME-BASED LIGHT-WEIGHT VIDEO CODING FOR MOBILE WIRELESS APPLICATION***Min Wu, Guogang Hua, Chang Wen Chen, Florida Institute of Technology, United States*

In conventional video coding, the complexity of an encoder is generally much higher than that of a decoder because of operations such as motion estimation consume significant computational resources. Such codec architecture is suitable for downlink transmission model of broadcast. However, in the contemporary applications of mobile wireless video uplink transmission, it is desirable to have low complexity video encoder to meet the resource limitations on the mobile devices. Recent advances in distributed video source coding provide potential reverse in computational complexity for encoder and decoder [cite{prism, aaron}]. In the same spirit, we proposed in this paper a syndrome-based light-weight video encoding scheme for mobile wireless applications. This scheme is based on two innovations: (1) adoption of low resolution low quality reference frames for motion estimation at the decoder; (2) introduction of more powerful product accumulate code. Extensive experimental results have confirmed that this syndrome based encoding can reduce computational complexity at the encoder while maintaining good reconstruction quality at the decoder. Therefore, this light weight video coding scheme is suitable for mobile wireless applications.

11:50 - 12:10

**WA2-L6.3 A FAST DOWNSIZING VIDEO TRANSCODER FOR H.264/AVC WITH RATE-DISTORTION OPTIMAL MODE DECISION***Huifeng Shen, University of Science and Technology of China, China; Xiaoyan Sun, Feng Wu, Microsoft Research Asia, China; Houqiang Li, University of Science and Technology of China, China; Shipeng Li, Microsoft Research Asia, China*

This paper focuses on the mode decision and motion selection problem when H.264/AVC video streams are transcoded in spatial resolution. A fast downsizing transcoding scheme is developed in which a new rate-distortion (R-D) optimal mode decision mechanism is presented for high speed transcoding as well as high coding efficiency. A model for estimating relative prediction errors is applied in this paper, which is free from computation of interpolation and SAD/SSD computation. Based on the selected model, a motion refinement within distance of 1 pixel is performed after mode decision. Experimental results demonstrate that our method can significantly speed up the spatial resolution reduction process, while achieving high coding efficiency.

## **WA2-L6.4 EFFECTIVE ERROR CONCEALMENT ALGORITHM BY BOUNDARY INFORMATION FOR H.264 VIDEO DECODER**

*Guang-Tung Jian, Mei-Juan Chen, Ming-Chieh Chi, National Dong Hwa University, Taiwan*

H.264 is a new video coding standard which contains some novel coding functions and provides more coding efficiency than previous standards. However, video streams are sensitive to channel errors that often degrade the visual quality of the decoded sequence, not only corrupt the current frame but also propagate to subsequent frames. In order to solve this problem, the error concealment in the decoder is very important and useful, neither bit rate nor delay will increase during the transmission. In this paper, we propose an error concealment method by utilizing the temporal correlation of the lost macroblock and its neighboring macroblocks to reconstruct the damaged data. Simulation results show that both of the subjective and objective performances of the proposed techniques are superior to conventional temporal concealment methods for H.264 standard.

### **WP1-L1: Wednesday, July 12, 13:40 - 15:00**

#### **WP1-L1 Mobile Multimedia Applications I (Lecture)**

Time: Wednesday, July 12, 13:40 - 15:00

Place: Carmichael

Chair: Mihaela van der Schaar, University of California, Los Angeles

13:40 - 14:00

#### **WP1-L1.1 AN EMBEDDED SUGGESTIVE INTERFACE FOR MAKING HOME VIDEOS**

*Brett Adams, Svetha Venkatesh, Curtin University of Technology, Australia*

This paper describes a novel suggestive interface embedded in a smart camera prototype aimed at aiding home movie makers. We focus on the problem of generating shot capture suggestions suitable to the user's filming context, intended audience and style, and formulate a novel aesthetic measure by which to judge proposed suggestions. Tight coupling between media and software allows the aesthetic measure to be sensitive to previous footage captures, including those taken without the system's prompting, in a manner allowing flexible, end-to-end migration of the authoring task from user to machine. An approximate method is used to find timely, near-optimal solutions to the aesthetic measure. Qualitative evaluation in the form of a user study shows it to be a promising approach to the flexible home movie authoring context.

14:00 - 14:20

#### **WP1-L1.2 MORE: A MOBILE OPEN RICH MEDIA ENVIRONMENT**

*Vidya Setlur, Tolga Capin, Suresh Chitturi, Nokia Research Center, United States; Ramakrishna Vedantham, Nokia, United States; Michael Ingrassia, Nokia Research Center, United States*

'Rich media' is a term that implies the integration of all of the advances we have made in the mobile space delivering music, speech, text, graphics and video. This is true, but it is more than the sum of its parts. Rich media is the ability to deliver these modalities, to interact with these modalities, and to do it in a way that allows for the construction, delivery and use of compelling mobile services in an effective and economic manner. In this paper, we introduce a system called Mobile Open Rich-media Environment ('MORE') that helps realize such mobile rich media services, combining various technologies of W3C, OMA, 3GPP and IETF standards. The different components of the system include formatting, packaging, transporting, rendering and interacting with rich media files and streams.

14:20 - 14:40

### **WP1-L1.3 TOWARDS AUTOMATIC MOBILE BLOGGING**

*Pujianto Cemerlang, Nanyang Technological University, Singapore; Joo-Hwee Lim, Yilun You, Institute for Infocomm Research, Singapore; Jun Zhang, Nanyang Technological University, Singapore; Jean-Pierre Chevallet, Institute for Infocomm Research, Singapore*

Weblog has gained its popularity lately. There are about 70,000 new blogs a day and about 29,100 blog updates an hour. With the blogging phenomenon and the proliferation of camera phones, mobile bloggers can write their blogs almost instantaneously. But how much further can current mobile blogging tools enhance the experience? In this paper, we propose a Mobilog framework to automate context-relevant annotation and synthesise personalised content for mobile blogging. In particular, we describe a system implementation of the framework, Travelog, adapted for tourism applications. Finally, we discuss the challenges and future possibilities for mobile blogging.

14:40 - 15:00

### **WP1-L1.4 MODELING INTERACTIONS FROM EMAIL COMMUNICATION**

*Dong Zhang, Daniel Gatica-Perez, IDIAP Research Institute, Switzerland; Deb Roy, MIT Media Lab, United States; Samy Bengio, IDIAP Research Institute, Switzerland*

Email plays an important role as a medium for the spread of information, ideas, and influence among its users. We present a framework to learn topic-based interactions between pairs of email users, i.e., the extent to which the email topic dynamics of one user are likely to be affected by the others. The proposed framework is built on the influence model and the probabilistic latent semantic analysis (PLSA) language model. This paper makes two contributions. First, we model interactions between email users using the semantic content of email body, instead of email header. Second, our framework models not only email topic dynamics of individual email users, but also the interactions within a group of individuals. Experiments on the Enron email corpus show some interesting results that are potentially useful to discover the hierarchy of the Enron organization.

## **WP1-L2: Wednesday, July 12, 13:40 - 15:00**

### **WP1-L2 Programmable Hardware Platform for Multimedia Systems and Applications (Lecture)**

Time: Wednesday, July 12, 13:40 - 15:00

Place: Johnston

Chair: Mohanarajah Sinnathamby, ViXS Systems Inc.

13:40 - 14:00

### **WP1-L2.1 MULTI-ISSUE MULTI-THREADED STREAM PROCESSOR**

*Somayeh Sardashti, Hamid Reza Ghasemi, Omid Fatemi, University of Tehran, Iran*

The MISP Processor is a programmable media processor which supports multi-issuing, multi-threading and stream processing techniques. MISP executes applications that have been mapped to the stream programming model. The stream model decomposes applications into a set of computation kernels that operate on data streams. This mapping exposes the inherent locality and parallelisms in media applications. MISP exploits thread level (TLP), data level (DLP), sub-word (SP) and instruction level (ILP) parallelisms inherent in media applications. Full simulator of MISP has been implemented and several media workloads composed of EEMBC benchmarks have been applied [1]. Also we applied test programs on Imagine stream processor [2]. The simulation results show MISP gains IPC of more than 2.08 times and performance of more than 1.86 times over Imagine. The synthesis results show area overhead per thread addition in MISP is about 7% without changing clock frequency.



14:00 - 14:20

**WP1-L2.2 ENHANCED ARCHITECTURAL SUPPORT FOR VARIABLE-LENGTH DECODING**

*Mohanarajah Sinnathamby, ViXS Systems, Canada; Subramania Sudharsanan, Naraig Manjikian, Queen's University, Canada*

This paper proposes a new architecture for efficient variable-length decoding (VLD) of entropy-coded data for multimedia applications on general-purpose processors. It enhances earlier proposals for low-complexity performance-enhancing hardware structures that exploit prefix/suffix properties of variable-length codes for common multimedia formats. The enhanced architecture is compared to previous architectures in terms of complexity and speed for FPGA implementation, and also in terms of area requirements, power consumption, and speed for a 0.18um ASIC fabrication process. Simulation results are reported for a pipelined processor with caches executing MPEG-4 software where VLD performance is doubled by incorporating the proposed architecture.

14:20 - 14:40

**WP1-L2.3 MOTION ESTIMATION FOR H.264/AVC USING PROGRAMMABLE GRAPHICS HARDWARE**

*Chi-Wang Ho, Oscar C. Au, S.-H. Gary Chan, Shu-Kei Yip, Hoi-Ming Wong, Hong Kong University of Science and Technology, Hong Kong SAR of China*

We present an efficient implementation of motion estimation (ME) for H.264/AVC using programmable graphics hardware. The cost function for ME in H.264/AVC depends on the motion vector (MV) predictor which is the median MV of three neighboring coded blocks. Previous implementations assume no dependency among adjacent blocks, which is not true for H.264/AVC, they also perform unsatisfactorily because of their low arithmetic intensity, which is defined as operation per word transferred. To overcome the dependency problem, we introduce a new implementation which performs ME on block-by-block basis. Moreover, we can adjust the arithmetic intensity easily to optimize the performance on different graphics cards. Experimental results show that our implementation is substantially faster (by 10 times) than our SIMD optimized CPU implementation.

14:40 - 15:00

**WP1-L2.4 ARCHITECTURE ANALYSIS FOR LOW-DELAY VIDEO CODING**

*Ralf M. Schreier, A. M. Tushar Iqbal Rahman, Ganesh Krishnamurthy, Albrecht Rothmel, University of Ulm, Germany*

Low-delay video coding is a key technology for video conferencing as well as upcoming remote-monitoring and automotive video applications like rear-view cameras or night vision systems. As the ongoing progress in programmable DSP and ASIC technology allows cost effective and flexible implementations of the necessary hardware, compressed video transmission systems over multimedia busses will soon replace the current uncompressed systems even in latency critical applications. In this paper, fundamentals and theoretic limits of low-delay video coding are discussed with respect to architectural consequences of real-time implementations. A general latency analysis for a compressed video transmission systems is presented considering algorithmic, architectural and transmission related delays.

**WP1-L3 Stereo, Multiview and 3D (Lecture)**

Time: Wednesday, July 12, 13:40 - 15:00

Place: Thomson

Chair: Gerasimos Potamianos, IBM - T.J. Watson Research Center

13:40 - 14:00

**WP1-L3.1 A COMPARISON OF THREE 3-D FACIAL RECONSTRUCTION APPROACHES**

*Alexander Woodward, Da An, Georgy Gimel'farb, Patrice Delmas, University of Auckland, New Zealand*

We compare three Computer Vision approaches to 3-D reconstruction, namely passive Binocular Stereo and active Structured Lighting and Photometric Stereo, in application to human face reconstruction for modelling virtual humans. An integrated lab environment was set up to simultaneously acquire images for 3-D reconstruction and corresponding data from a 3-D scanner. This allowed us to quantitatively compare reconstruction results to accurate ground truth. Our goal was to determine whether any current Computer Vision approach is accurate enough for practically useful 3-D facial surface reconstruction. Comparative experiments show the combination of Structured Lighting with Symmetric Dynamic Programming based Binocular Stereo has good prospects due to reasonable processing time and sufficient accuracy.

14:00 - 14:20

**WP1-L3.2 PERSON TRACKING IN SMART ROOMS USING DYNAMIC PROGRAMMING AND ADAPTIVE SUBSPACE LEARNING**

*Zhenqiu Zhang, Beckman Institute, United States; Gerasimos Potamianos, Stephen Chu, IBM T. J. Watson Research Center, United States; Jilin Tu, Beckman Institute, United States; Thomas S. Huang, University of Illinois at Urbana-Champaign, United States*

We present a robust vision system for single person tracking inside a smart room using multiple calibrated cameras. The system consists of two components, initialization and tracking, assisted by tracking drift detection. The main novelty lies in the adaptive tracking mechanism, based on subspace learning of the tracked person appearance in selected camera views, with an additional "forgetting" mechanism introduced to reduce drifting. By combining the proposed technique with a robust initialization component, based on face detection and spatio-temporal dynamic programming, the resulting vision system significantly outperforms previously reported systems on data collected as part of the CHIL project.

14:20 - 14:40

**WP1-L3.3 AN EFFICIENT LOCAL CLUSTERING APPROACH FOR SIMPLIFICATION OF 3D POINT-BASED COMPUTER GRAPHICS MODELS**

*ZhiWen Yu, Hau-San Wong, City University of Hong Kong, Hong Kong SAR of China*

Given a point-based 3D computer graphics model which is defined by a point set  $P$  and a desired reduced number of output samples  $N_s$ , the simplification approach finds a point set  $P_s$  which (i) satisfies  $|P_s| = N_s$  ( $|P_s|$  is the cardinality of  $P_s$ ) and (ii) minimizes the difference of the corresponding surface  $S_s$  (defined by  $P_s$ ) and the original surface  $S$  (defined by  $P$ ). Although a number of previous approaches have been proposed for simplification, most of them (i) do not focus on pointbased 3D models, (ii) do not consider efficiency, quality and generality together. In this paper, we introduce an adaptive simplification method (ASM) which is an efficient technique for simplifying point-based complex 3D model. ASM achieves low running time by clustering the points locally based on the preservation of geometric characteristics. Finally, we analyze the performance of ASM and show that it outperforms most of the current state-of-the-art methods in terms of efficiency, quality and generality.

**WP1-L3.4 REAL-TIME DEPTH IMAGE BASED RENDERING HARDWARE ACCELERATOR FOR ADVANCED THREE DIMENSIONAL TELEVISION SYSTEM**

*Wan-Yu Chen, Yu-Lin Chang, Hsu-Kuang Chiu, Shao-Yi Chien, Liang-Gee Chen, National Taiwan University, Taiwan*

3D TV will become a prominent technology in the next generation. In this paper, a depth image based rendering system is proposed from algorithm level to hardware architecture level. We propose a novel depth image based rendering algorithm with edge-dependent gaussian filter and interpolation to improve the rendered stereo image quality. Based on our proposed algorithm, a fully-pipelined depth image based rendering hardware accelerator is proposed to support real-time rendering. The proposed hardware accelerator is optimized in three steps. First, we analyze the effect of fixed point operation and choose the optimal wordlength to keep the stereo image quality. Second, a three-parallel edge-dependent gaussian filter architecture is proposed to solve the critical problem of memory bandwidth. Finally, we optimize the hardware cost by the proposed hardware architecture. Only 1/21 amounts of vertical PEs and 1/11 amounts of horizontal PEs is needed by the proposed folded edge-dependent gaussian filter architecture. Furthermore, by the proposed check mode, the whole Z-buffer can be eliminated during 3D image warping. In additions, the on-chip SRAMs can be reduced to 66.7 percent compared with direct implementation by global and local disparity separation scheme. A prototype chip can achieve real-time requirement under the operating frequency of 80 MHz for 25 SDTV frames per second (fps) in left and right channel simultaneously. The simulation result also shows the hardware cost is quite small compared with the conventional rendering architecture.

### WP1-L4: Wednesday, July 12, 13:40 - 15:00

**WP1-L4 Multimedia Technologies for Distance Learning (Special Session)**

Time: Wednesday, July 12, 13:40 - 15:00

Place: Toronto I

Co-Chairs: Qing Li, City University of Hong Kong and Timothy K. Shih, Tamkang University, Taiwan

13:40 - 13:53

**WP1-L4.1 WEB BASED CHINESE CALLIGRAPHY LEARNING WITH 3-D VISUALIZATION METHOD**

*Yingfei Wu, Yueting Zhuang, Yunhe Pan, Jiangqin Wu, Zhejiang University, China*

Chinese calligraphy is pictographic and each calligraphist has his own writing style. People often feel difficult in writing a demanded beautiful calligraphy style. In order to help people enjoy the art of calligraphy and learn how it is written step-by-step we present a new approach to animate its writing process by 3-D visualization method. In this paper some novel algorithms used in the approach are presented to solve the following problems: 1) estimate varied stroke's thickness 2) extract strokes order from an offline Chinese calligraphic writing. Through this approach we implement a system. Experimental result is given to demonstrate the application finally.

13:53 - 14:06

**WP1-L4.2 TOWARDS THE APPLICATION OF A HANDWRITING INTERFACE FOR MATHEMATICS LEARNING**

*Lisa Anthony, Jie Yang, Kenneth R. Koedinger, Human-Computer Interaction Institute, United States*

We believe handwriting input may be able to provide significant advantages over typing, especially in the mathematics learning domain. The use of handwriting may result in decreased extraneous cognitive load on students, and it may provide better support for the two-dimensional spatial components of mathematics when compared to existing typing-based tools. Here we report progress towards the application of a handwriting interface for mathematics learning. We introduce a prototype system that allows students to use handwriting input to solve algebraic equations in an intelligent tutor. We discuss strategies to improve the existing handwriting system and apply it to math learning. Although the recognition accuracy of current handwriting engines may not be at a level suitable for use by students, we hypothesize that this may be realistically improved via advance training of the engine on a large corpus, as well as via techniques similar to co-training.

14:06 - 14:19

**WP1-L4.3 A VIOLIN MUSIC TRANSCRIBER FOR PERSONALIZED LEARNING**

*Wei Jie Jonathan Boo, Ye Wang, Alex Loscos, National University of Singapore, Singapore*

This paper presents a new version of our violin music transcriber [1] to support personalized learning. The proposed method is designed to detect duo-pitch (two strings being bowed at the same time) from real-world violin audio signals recorded in a home environment. Our method uses a semitone band spectrogram, a signal spectral representation with direct musical relevance. We exploit constraints of violin sound to improve the transcription performance and speed in comparison with existing methods. We have carried out rigorous evaluations using (a) single pitch notes and duo-phonetic pitch samples within the violin's playing range (G3-B6), and (b) music excerpts. For pitch and duo-pitch samples our method can achieve a transcription precision score of 93.1% and recall score of 96.7% respectively. For music excerpts, an average of 95% of all notes could be found (recall), and 93% of notes transcribed correctly (precision).

14:20 - 14:33

**WP1-L4.4 PERSONALIZED SEARCH OF EDUCATIONAL CONTENT BASED ON MULTIPLE ONTOLOGIES**

*Apple W. P. Fok, Horace H. S. Ip, City University of Hong Kong, Hong Kong SAR of China*

In this paper, we introduce the design of Personalized Education (PE) search approach that employs multiple ontologies to automatically generate queries for educational resources retrieval based only a high level specification of the teaching/learning needs of a user. Central to this approach is the design of PEOnto which is an educational ontology that consists of FIVE interrelated educational ontologies that supports the delivery of personalized educational services in a Personalized Education System. We illustrate the feasibility of PEOnto design through a scenario walkthrough in the context of Grade 4 English Language learning.

14:33 - 14:46

**WP1-L4.5 ADAPTIVE MULTIMEDIA FLOW REGULATION AND DELAY ANALYSIS FOR END-HOSTS IN P2P NETWORKS**

*Weijia Jia, City University of Hong Kong, Hong Kong SAR of China*

This paper proposes a novel adaptive algorithm based on a new soft-regulator ( $\sigma, \rho, \lambda$ ) for concurrent multimedia flows at end hosts. Our algorithm has the following features: (1) does requires the support of the network layers (routers and switches); (2) is scalable as it can be installed into any intermediated nodes for the delay control; (3) uses network resource efficiently, in particular, when the bandwidth of network is very limited. Performance experimental data have shown that ( $\sigma, \rho, \lambda$ ) and related algorithm are efficient and applicable as compared with network layer solutions.

14:46 - 14:59

**WP1-L4.6 ERMDCLIME: ENABLING REAL-TIME MULTIMEDIA DISCUSSION FOR COLLABORATIVE LEARNING IN MOBILE ENVIRONMENT**

*Xin Xiao, Yuan-Chun Shi, Nan Zhang, Tsinghua University, China*

With the aid of mobile technology, it's possible to carry out collaborative learning (CL) in any places at any time. And the most essential and effective way to achieve this goal is through real-time multimedia discussion. In this paper, a multi-agent system, ErmdClime, is designed and implemented to enable real-time multimedia discussion for collaborative learning in mobile environment. The real-time bi-directional audio/video interaction and shared white board are integrated seamlessly into the system for learners with mobile devices. Besides, this system enables both local and remote learners to control the play of PowerPoint slides by voice command when discussing and annotate on the shared white board. Context-aware technology is used to switch real video and white board display on mobile devices. User study demonstrates that learners feel quite natural and highly efficient when they are engaged in the CL with mobile devices in our system.

**WP1-L5 Feature Extraction and Segmentation IV (Lecture)**

Time: Wednesday, July 12, 13:40 - 15:00

Place: Toronto II

Chair: Konstantinos Plataniotis, University of Toronto

13:40 - 14:00

**WP1-L5.1 HUMAN OBJECT TRACKING ALGORITHM WITH HUMAN COLOR STRUCTURE DESCRIPTOR FOR VIDEO SURVEILLANCE SYSTEMS***Shao-Yi Chien, Wei-Kai Chan, Der-Chun Cherng, Jing-Ying Chang, National Taiwan University, Taiwan*

Segmentation, tracking, and description extraction are important operations in smart camera surveillance systems. In this paper, a robust segmentation-and-descriptor based tracking algorithm is proposed. Segmentation is applied first, and description for each connected component is extracted for object classification to generate the video object masks. It can do segmentation, tracking, and description extraction with a single algorithm without redundant computation. In addition, a new descriptor for human objects, Human Color Structure Descriptor (HCSD), is also proposed for this algorithm. Experimental results show that the proposed algorithm can provide precise video object masks and trajectories. It is also shown that the proposed descriptor, HCSD, can achieve better performance than Scalable Color Descriptor and Color Structure Descriptor of MPEG-7 for human objects.

14:00 - 14:20

**WP1-L5.2 STORY BOUNDARY DETECTION IN NEWS VIDEO USING GLOBAL RULE INDUCTION TECHNIQUE***Lekha Chaisorn, Institute for Infocomm Research, Singapore; Tat-Seng Chua, National University of Singapore, Singapore*

Global rule induction technique has been successfully used in information extraction (IE) from text documents. In this paper, we employ the technique to identify story boundaries in news video. We divide our framework into two levels: shot and story levels. We use a hybrid algorithm to classify each input video shot into one of the predefined genre types and employ the global rule induction technique to extract story boundaries from the sequence of classified shots. We evaluate our rule induction based system on ~120-hours of news video provided by TRECVID 2003. The results show that we could achieve an F1 accuracy of over 75%.

14:20 - 14:40

**WP1-L5.3 APPLYING SUPERVISED CLASSIFIERS BASED ON NON-NEGATIVE MATRIX FACTORIZATION TO MUSICAL INSTRUMENT CLASSIFICATION***Emmanouil Benetos, Margarita Kotti, Constantine Kotropoulos, Aristotle University of Thessaloniki, Greece*

In this paper, a new approach for automatic audio classification using non-negative matrix factorization (NMF) is presented. Training is performed onto each audio class individually, whilst during the test phase each test recording is projected onto the several training matrices. Experiments demonstrating the efficiency of the proposed approach were performed for musical instrument classification. Several perceptual features as well as MPEG-7 descriptors were measured for 300 sound recordings consisting of 6 different musical instrument classes. Subsets of the feature set were selected using branch-and-bound search, in order to obtain the most discriminating features for classification. Several NMF techniques were utilized, namely the standard NMF method, the local NMF, and the sparse NMF. The experiments demonstrate an almost perfect classification (classification error 1.0%), outperforming the state-of-the-art techniques tested for the aforementioned experiment.

14:40 - 15:00

#### **WP1-L5.4 SCALABILITY IN HUMAN SHAPE ANALYSIS**

*Thomas Foures, Philippe Joly, IRIT, France*

This paper proposes a new approach for the human motion analysis. The main contribution comes from the proposed representation of the human body. Most of already existing systems are based on a model. When this one is a priori known, it may not evolve automatically according to user needs, or to the detail level that is actually possible to extract, or to restrictions due to the processing time. In order to propose a more flexible system, a hierarchical representation of the human body is implemented. It aims at providing a multi-resolution description and results at different levels of accuracy. An explanation about the model construction and the method used to map it onto features extracted from an image sequence are presented. Relations between the different body limbs and some physical constraints are then integrated. The transition from a model level to the next one is also explained and results on frames coming from a video sequence give an illustration of the proposed strategy.

### **WP1-L6: Wednesday, July 12, 13:40 - 15:00**

#### **WP1-L6 Multimedia for Virtual Reality (Lecture)**

Time: Wednesday, July 12, 13:40 - 15:00

Place: Varley

Chair: Toshio Asano, Nagoya Institute of Technology

13:40 - 14:00

#### **WP1-L6.1 INTERACTIVE HAPTIC TRANSMISSION FOR REMOTE CONTROL SYSTEMS**

*Toshio Asano, Yutaka Ishibashi, Seiji Kameyama, Nagoya Institute of Technology, Japan*

In this paper, we propose a remote control system which controls a haptic interface device with another remote haptic interface device. Haptic feeling can be transmitted by the system. Applications of the system are a remote calligraphy system, a remote drawing instruction system, a remote medical operation system, and so on. The paper deals with the remote drawing instruction system by which an instructor trains a learner how to draw pictures or figures while conveying the sense of force through the Internet. In order to clarify what kinds of control are needed in the system, we examine the influences of network delay, delay jitter and packet loss on the output quality of haptic media by subjective assessment of drawing figures.

14:00 - 14:20

#### **WP1-L6.2 REMOTE CONTROL OF A NETWORKED MOBILE ROBOT USING AN IMMERSIVE LOCOMOTION INTERFACE**

*Kazumasa Yamazawa, Masaki Tawada, Naokazu Yokoya, Nara Institute of Science and Technology, Japan*

Recently, the concept of behavioral media using a mobile robot has emerged in tele-robotics, where the mobile robot behaves as a user's second body or an agent in a remote site. An important issue in behavioral media is how to control the robot intuitively immersing a user into a remote site. This paper describes a new approach which combines locomotion interface and immersive projection display of a remote site for controlling a networked mobile robot. We have developed a prototype system and have evaluated the usability of the system with experiments.

14:20 - 14:40

**WP1-L6.3 MOVIES MADE THROUGH THE EYES OF A MOBILE USER WITH A GAZE-ALIGNED CAMERA**

*Erich Schneider, Klaus Bartl, Guido Böning, Thomas Dera, University of Munich, Germany; Philipp Wagner, Technical University of Munich, Germany; Stanislavs Bardins, University of Munich, Germany; Mitch Enzmann, 40° Filmproduktion GmbH, Germany; Thomas Brandt, University of Munich, Germany*

Head-fixed camera systems are widely known, but since they are aligned by the head only and not by the eyes they are not able to always “look” at what the cameraman is looking at and the image quality is poor if no effort is made to stabilize the camera. Systems like Steadycam, in contrast, focus on image stabilization at the cost of restricting the cameraman in his actions. The prototype of a new mobile head-mounted camera system was developed that is continuously aligned with the orientation of gaze. In doing so, the biological gaze stabilization reflexes are used to keep the video camera stable on target. Applications like movie making through the eyes of an actor or documentary movies for sports and other activities are conceivable. The system was tested by a surgeon who could successfully document his activities.

14:40 - 15:00

**WP1-L6.4 GPU-BASED OCEAN RENDERING**

*Yung-Feng Chiu, Chun-Fa Chang, National Tsing Hua University, Taiwan*

We present efficient algorithms for real-time rendering of ocean using the newest features of programmable graphics processors (GPU). It differs from previous works in three aspects: adaptive GPU-based ocean surface tessellation, sophisticated optical effects for shallow water, and spray dynamics for oscillating waves. Our tessellation scheme not only offers easier level-of-detail (LOD) control but also avoids the loading of vertex attributes from CPU to GPU at each frame. The object-space wave sampling approach allows us to produce sophisticated optical effects for shallow water and implement a state-preserving particle system for simulating spray motions interactively.

**WP2-L1: Wednesday, July 12, 15:30 - 16:50**

**WP2-L1 Mobile Multimedia Applications II (Lecture)**

Time: Wednesday, July 12, 15:30 - 16:50

Place: Carmichael

Chair: Hari Kalva, Florida Atlantic University

15:30 - 15:50

**WP2-L1.1 CMUSEUM: A LOCATION-AWARE WIRELESS VIDEO STREAMING SYSTEM**

*Mei-Hsuan Lu, Tsuhan Chen, Carnegie Mellon University, United States*

Location-aware computing enables automatic tailoring of information and services based on the current location of a mobile user. We have designed and implemented CMuseum, a system that enables location-based video streaming applications, as well as other add-on services on top of the purposed infrastructure. Our novel design incorporates an 802.15.4 Zigbee sensor network for collecting location information as well as an 802.11 Wi-Fi network for streaming video contents. This framework is largely event-driven in order to support the real-time nature of the video streaming services it handles. We have demonstrated the viability of this system through implementing a working system in a museum tour guide scenario.



15:50 - 16:10

**WP2-L1.2 COMPUTING A MULTIMEDIA REPRESENTATION FOR DOCUMENTS GIVEN TIME AND DISPLAY CONSTRAINTS**

*Berna Erol, Kathrin Berkner, Ricoh California Research Center, United States; Siddharth Joshi, Stanford University, United States; Jonathan J. Hull, Ricoh California Research Center, United States*

It is difficult to view multipage, high resolution documents on devices with small displays. As a solution, we introduce a Multimedia Thumbnail representation, which can be seen as a multimedia clip that provides an automated guided tour through a document. Multimedia Thumbnails are automatically generated by taking a document image as input and first performing visual and audible information analysis on the document to determine salient document elements. Next, the time and information attributes for each document element are computed by taking into account the display and application constraints. An optimization routine, given a time constraint, selects elements to be included in the Multimedia Thumbnail. Last, the selected elements are synthesized into animated images and audio to create the final multimedia representation.

16:10 - 16:30

**WP2-L1.3 DIRECTING ATTENTION IN ONLINE AGGREGATE SENSOR STREAMS VIA AUDITORY BLIND VALUE ASSIGNMENT**

*Robert Malkin, Datong Chen, Jie Yang, Alex Waibel, Carnegie Mellon University, United States*

In real-world multiparty collaborative applications, it is useful for a central planning agent to receive online audiovisual information from all participants. As the size of the group grows, it becomes difficult to process all the sensory streams; cognitive overload prevents analysis of sensory streams for situational awareness. To avoid this situation, an automatic method is needed to assign value to each stream and direct the planning agent's attention to the most valuable streams. We present an audio-based blind value assignment method to address this problem, and demonstrate the method's efficacy. Our experiments show that audio BVA produces value judgments which are similar to human judgments and superior to video-based judgments.

16:30 - 16:50

**WP2-L1.4 PERMUTATION-BASED LOW-COMPLEXITY ALTERNATE CODING IN MULTI-VIEW H.264/AVC**

*Daniel Socek, Dubravko Culibrk, Hari Kalva, Oge Marques, Borko Furht, Florida Atlantic University, United States*

Low bitrates enabled by the H.264/AVC standard come at the cost of significant decoding complexity. The problem is exacerbated when multi-view video coding (MVC), presently being standardized by the MPEG committee, is considered. Increasing the number of reference views will improve the prediction but also increase the encoding and decoding complexity. In this paper we present an alternative to transform coding and arithmetic coding used to code motion compensation residuals in H.264/AVC. Proposed approach is sorting permutation based, resulting in significantly decreased decoding complexity. We evaluate the efficiency of the coding and present initial results obtained for some sequences in published literature.

**WP2-L2 Protocols for Multimedia Delivery Over Networks (Lecture)**

Time: Wednesday, July 12, 15:30 - 16:50

Place: Johnston

Chair: Bo Shen, Hewlett-Packard Labs

15:30 - 15:50

**WP2-L2.1 A CASE FOR INTERNET STREAMING VIA WEB SERVERS**

*Songqing Chen, George Mason University, United States; Bo Shen, Wai-tian Tan, Susie Wee, HP Laboratories, United States; Xiaodong Zhang, The Ohio State University, United States*

Hosting Internet streaming services has its unique challenges. Aiming at making Internet streaming services be widely and easily adopted in practice, in this paper, we have designed and implemented a system, called SProxy that can leverage existing Internet infrastructure to free the streaming content providers so that they only need to host streaming content through a regular Web server. SProxy has been extensively tested and evaluated and it provides high quality streaming delivery in both local area networks and wide area networks (e.g. between Japan and US).

15:50 - 16:10

**WP2-L2.2 PASS: PEER-AWARE SILENCE SUPPRESSION FOR INTERNET VOICE CONFERENCES**

*Xun Xu, University of Illinois at Urbana-Champaign, United States; Li-wei He, Dinei Florêncio, Yong Rui, Microsoft Research, United States*

A novel tandem free solution for multiparty VoIP conferences called PASS (Peer-Aware Silence Suppression) is presented. Through introducing a limit on the number of concurrent speakers in a conference, PASS involves silence suppression and speaker selection which are completely distributed to run on each client. This configuration leads to better scalability, lower bandwidth occupation and jitter buffer delay, and higher compatibility with a wide variety of network topologies. The key component of PASS, distributed silence suppression and speaker selection, is realized through a robust approach proposed in this paper. Based on a voice activity measure derived using machine learning techniques, this approach is able to reliably suppress silence in complex environments, and perform accurate and transparent speaker selection as well.

16:10 - 16:30

**WP2-L2.3 ANALYSIS AND EVALUATION OF THE SKYPE AND GOOGLE-TALK VOIP SYSTEMS**

*Batu Sat, Benjamin W. Wah, University of Illinois at Urbana-Champaign, United States*

In this paper, we study Skype and Google Talk, two widely used VoIP systems, and compare their perceptual speech quality with that of our proposed system using UDP packet traces collected in the PlanetLab. Based on methods for speech coding, packetization, jitter control, estimation and feedback of network conditions, and loss concealments, our results show that Skype has noticeable quality degradations because it only uses a maximum of two-way redundancy for loss concealment, does not handle out-of-order arrivals, and applies a fixed jitter control of 60 ms relative to the expected arrival time. Its slow loss-adaptation time of more than one minute to change from one-way to two-way redundancy makes it susceptible to quality degradation under fast changing loss conditions. In contrast, Google Talk does not employ any loss adaptation and performs similar to Skype under low- to medium-loss scenarios. By addressing the shortcomings of Skype and Google Talk, we demonstrate improvements in speech quality in our proposed prototype.

16:30 - 16:50

**WP2-L2.4 RATE-DISTORTION OPTIMIZED VIDEO STREAMING FOR SCALABLE H.264**

*Sangho Yoon, Stanford university, United States; Mark Mao, Mark Kalman, Stanford University, United States*

We propose a new real-time packet scheduling algorithm for streaming scalable H.264. Our algorithm makes use of a packet importance measure, which we define, that takes into consideration transmission history, channel conditions, and the unique decoding dependencies due to the temporal wavelet encoding. Our algorithm utilizes this importance measure to minimize the expected reconstruction distortion at the decoder under a certain rate constraint. In our experimental results we show gains of more than 3 dB in decoded video quality when transmissions are controlled with our algorithm as compared to existing schedulers.

**WP2-L3 3D-TV: Primed for Success? III (Special Session)**

Time: Wednesday, July 12, 15:30 - 16:50

Place: Thomson

Co-Chairs: Liang Zhang, Communications Research Centre Canada, Wa James Tam, Communications Research Centre Canada and Hirokazu Yamanoue, NHK Science & Technical Research Laboratories

15:30 - 15:50

**WP2-L3.1 3D VIDEO AND FREE VIEWPOINT VIDEO – TECHNOLOGIES, APPLICATIONS AND MPEG STANDARDS**

*Aljoscha Smolic, Karsten Mueller, Philipp Merkle, Christoph Fehn, Peter Kauff, Peter Eisert, Thomas Wiegand, Fraunhofer – Heinrich Hertz Institute, Germany*

An overview of 3D and free viewpoint video is given in this paper with special focus on related standardization activities in MPEG. Free viewpoint video allows the user to freely navigate within real world visual scenes, as known from virtual worlds in computer graphics. Examples are shown, highlighting standards conform realization using MPEG-4. Then the principles of 3D video are introduced providing the user with a 3D depth impression of the observed scene. Example systems are described again focusing on their realization based on MPEG-4. Finally multi-view video coding is described as a key component for 3D and free viewpoint video systems. The conclusion is that the necessary technology including standard media formats for 3D and free viewpoint is available or will be available in the near future, and that there is a clear demand from industry and user side for such applications. 3DTV at home and free viewpoint video on DVD will be available soon, and will create huge new markets.

15:50 - 16:10

**WP2-L3.2 CARRIAGE OF 3D AUDIO-VISUAL SERVICES BY T-DMB**

*Sukhee Cho, Namho Hur, Jinwoong Kim, Kugjin Yun, Soo-In Lee, Electronics and Telecommunications Research Institute (ETRI), Republic of Korea*

In this paper, we introduce our experience on the development of a three-dimensional audio-visual(3D AV) service system based on the terrestrial digital multimedia broadcasting (T-DMB) system. 3D AV service is now much more feasible than before with the fast advancement of hardware technologies, especially 3D flat panel display, processors and memory. 3D AV service over DMB system is very attractive due to the facts that (1) glassless 3D viewing with small display is relatively easy to implement and more suitable to the single user environment like DMB, (2) DMB is a new media and thus has more flexibility in adding new services on the existing ones, (3) 3D AV handling capability of 3D DMB terminal has lots of potential to generate new types of services if it is added with other components like built-in stereo camera. In order to provide successful 3D DMB services over existing DMB system, we need to solve several issues like (1) guaranteeing backward compatibility with the T-DMB system, (2) minimizing the overhead on the transmitted bit-rate and the required processing power of the terminal, (3) providing good 3D depth perception without a noticeable eye strain. We propose a very efficient and backward compatible system architecture for the 3D DMB, and show how we can get better depth perception with the limited bit budget of the DMB system.

16:10 - 16:30

**WP2-L3.3 END-TO-END STEREOSCOPIC VIDEO STREAMING SYSTEM**

*Selen Pehlivan, Koç University, Turkey; Anil Aksay, Cagdas Bilen, Gozde Bozdagi Akar, Middle East Technical University, Turkey; M. Reha Civanlar, Koç University, Turkey*

Today, stereoscopic and multi-view video are among the popular research areas in the multimedia world. In this study, we have designed and built a platform consisting of stereo-view capturing, real-time transmission and display. At the display stage, end users view video in 3D by using polarized glasses. Multi-view video is compressed in an efficient way by using multi-view video coding techniques and streamed using standard real-time transport protocols. The entire system is built by modifying available open source systems whenever possible. Receiver can view the content of the video built from multiple channels as mono or stereo depending on its display and bandwidth capabilities.

16:30 - 16:50

**WP2-L3.4 A REAL TIME INTERACTIVE DYNAMIC LIGHT FIELD TRANSMISSION SYSTEM**

*Yebin Liu, Qionghai Dai, Wenli Xu, Tsinghua University, China*

The ability to interactively and seamlessly roam in the scenario while watching a video through IP network is an exciting visual experience. In this work, we implemented a 3D TV system with real-time data acquisition, compression, internet transmission, light field rendering, and free-viewpoint control of dynamic scenes. Our system consists of an 8×8 light field camera array, 16 producer PCs, a streaming server system and several clients. Multiple video streams are coded in a real time manner that each client can freely select the streams for novel view rendering. Also, our system minimize the per-user transmission bit rate while maintaining multi-view simul-switching ability for each user. We believe that this is the first real-time internet streaming system that can simultaneously guarantee real time free-view point control, data storage and support arbitrary number of users. The average transmission bit rate for end user is lower than 2Mbps which is suitable for the broadband IP network.

**WP2-L4: Wednesday, July 12, 15:30 - 16:50**

**WP2-L4 Encryption and Access Control (Lecture)**

Time: Wednesday, July 12, 15:30 - 16:50

Place: Toronto II

Chair: Oscar Au, Hong Kong University of Science & Technology

15:30 - 15:50

**WP2-L4.1 LIGHTWEIGHT JPEG2000 ENCRYPTION WITH ANISOTROPIC WAVELET PACKETS**

*Dominik Engel, Andreas Uhl, University of Salzburg, Austria*

A lightweight encryption technique for JPEG2000 with optional support for transparent encryption is proposed. Randomized anisotropic wavelet packet bases are used to construct a secret frequency domain, leading to a situation in which only a minimal amount of data needs to be encrypted. Results and calculations are presented to evaluate the suggested approach in terms of compression performance, security, and applicability.

15:50 - 16:10

**WP2-L4.2 SECURED MPEG-21 DIGITAL ITEM ADAPTATION FOR H.264 VIDEO**

*Razib Iqbal, Shervin Shirmohammadi, Abdulmotaleb El Saddik, University of Ottawa, Canada*

Seamless adaptation and transcoding techniques to adapt the digital content have achieved significant focus to serve the consumers with the desired content in a feasible way. With the succession of time we sense that secured adaptation should also be taken care of for not only serving sensitive digital contents but also to offer security as an embedded feature of the adaptation practice to ensure digital right management and confidentiality. In this paper, we propose an encryption framework for a transcoder while adapting H.264 video conforming to MPEG-21 DIA. Encryption mechanism is applied on the adapted video content thus reducing computational overhead compared to that on the original content.

16:10 - 16:30

**WP2-L4.3 ON THE SECURITY OF MULTIMEDIA ENCRYPTION SCHEMES BASED ON MULTIPLE HUFFMAN TABLE (MHT)**

*Jiantao Zhou, Zhiqin Liang, Yan Chen, Oscar C. Au, Hong Kong University of Science and Technology, China*

This paper addresses the security issues of the multimedia encryption schemes based on multiple Huffman table (MHT). A detailed analysis of known-plaintext attack is presented to show that the Huffman tables used for encryption should be carefully selected to avoid the weak keys problem. Further, we propose an efficient chosen-plaintext attack on the basic MHT method as well as the enhanced scheme inserting random bits. We also show that random rotation in partitioned bit stream cannot essentially improve the security.

# AUTHOR INDEX

Aarabi, Parham .....	53, 137	Baraldi, Stefano .....	50
Abdel-Mottaleb, Mohamed .....	94	Barbieri, Mauro .....	134
Abe, Hironobu .....	114	Bardins, Stanislavs .....	170
Abed-Meraim, Karim .....	98	Barkat, Braham .....	98
Aboutajdine, Driss .....	44	Barkowsky, Marcus .....	32
Acero, Alex .....	105	Bartl, Klaus .....	170
Adams, Brett .....	162	Bashir, Faisal .....	51
Ahmad, Shakeel .....	42	Basso, Andrea .....	144
Ahn, ChungHyun .....	116	Basu, Anup .....	110
Aizawa, Kiyoharu .....	29, 154	Basu, Sumit .....	40, 120
Akon, Mohammad .....	81	Bates, Richard .....	140
Aksay, Anil .....	173	Bauchhage, Christian .....	54
Al-Hames, Marc .....	46	Bauer, Claus .....	75
Al-khassaweneh, Mahmood .....	131, 133	Baykal, Buyurman .....	91
Alattar, Ashraf .....	46	Bazzi, Rida .....	62
Ali, Sarah .....	137	Beauregard, Gerald .....	44
Almpanidis, George .....	92	Bellas, Nikolaos .....	112
Alusse, Yannick .....	67	Bellini, Pierfrancesco .....	57, 67
Alwan, Abeer .....	106	Ben-Jamaa, Salma .....	124
Amer, Aishy .....	46	Benetos, Emmanouil .....	101, 168
Amini, Lisa .....	93	Bengio, Samy .....	163
Amit, Guy .....	30	Benini, Sergio .....	93
Amores, Jaime .....	64	Bennour, Jihane .....	102
An, Da .....	165	Ben Amar, Chokri .....	132
Ansari, A-Nasser .....	94	Ben Hamadou, Abdelmajid .....	143
Antani, Sameer .....	160	Ben Hamza, Abdessamad .....	132
Anthony, Lisa .....	166	Berkner, Kathrin .....	171
Antini, Gianni .....	35	Bernstein, Lynne .....	106
Apostolopoulos, John .....	64, 85	Berretti, Stefano .....	35
Appel, Ron .....	53	Bertini, Marco .....	64
Aran, Oya .....	87	Bialkowski, Jens .....	32
Ardabilain, Mohsen .....	143	Bianchetti, Aldo .....	93
Argyriou, Antonios .....	106, 129	Bilen, Cagdas .....	173
Argyropoulos, Savvas .....	68	Bin, Li .....	122
Ariki, Yasuo .....	114	Bishop, William .....	75
Arsic, Dejan .....	100	Blanc-Talon, Jacques .....	101, 119
Arsic, Ivana .....	40	Böning, Guido .....	170
Asahara, Akihiro .....	57	Boo, Wei Jie Jonathan .....	167
Asano, Toshio .....	169	Borba, Gustavo .....	78
Atienza, David .....	98	Borgstrom, Jonas .....	106
Atrey, Pradeep Kumar .....	147	Boston, Nigel .....	143
Atzpadin, Nicole .....	152	Bouaziz, Bassem .....	143
Au, Oscar C. ....	81, 83, 84, 85, 164, 174	Bouguet, Jean-Yves .....	63
Avessta, Nastooh .....	75, 112	Boulgouris, Nikolaos V. ....	68
Aviyente, Selin .....	131, 133	Bouridane, Ahmed .....	159
Azuma, Shozo .....	53	Boutell, Matthew .....	79
Babaguchi, Noboru .....	87	Boyer, Jean-Philippe .....	101, 119
Bambos, Nick .....	85	Bozdagi Akar, Gozde .....	173

Brandt, Thomas .....	170	Chen, Ching-Yeh .....	52
Brecheisen, Stefan .....	120	Chen, Chu-Song .....	77, 97
Brown, Christopher .....	79	Chen, Chun .....	137
Bruno, Ivan .....	57	Chen, Datong .....	95, 160, 171
Bu, Jiajun .....	137	Chen, Feng .....	96
Burnett, Ian .....	128	Chen, Gao .....	50
Byun, Hyung-Gi .....	115	Chen, Jia-Wei .....	99
Cai, Hua .....	110	Chen, Jianfeng .....	156
Cai, Jianfei .....	112, 138, 150	Chen, Jyi-Feng .....	41
Candan, K. Selçuk .....	66	Chen, Lei .....	79
Cao, Gang .....	50	Chen, Liang-Gee .....	43, 47, 52, 156, 166
Cao, Wanyong .....	125	Chen, Mei-Juan .....	162
Capin, Tolga .....	162	Chen, Meng .....	104
Cardoso, Jaime .....	101	Chen, Ming-Syan .....	104, 156
Carvalho, Paulo .....	133	Chen, Ming-yu .....	48
Casetti, Claudio .....	90	Chen, Siyue .....	113
Casey, Edward .....	69	Chen, Songqing .....	172
Caspi, Yaron .....	30	Chen, Tien-Fu .....	30
Catthoor, Francky .....	98	Chen, Tsuhan .....	35, 170
Cemerlang, Pujianto .....	163	Chen, Tung-Chien .....	47, 156
Chaisorn, Lekha .....	168	Chen, Wan-Yu .....	166
Chakareski, Jacob .....	128, 129	Chen, Wei-Ta .....	156
Chalhoub, Georges .....	54	Chen, Wen-Feng .....	47
Chan, Carri .....	85	Chen, Xiexiong .....	42
Chan, H. Anthony .....	114	Chen, Xilin .....	56
Chan, Raymond K.W. ....	155	Chen, Xin .....	90
Chan, S.-H. Gary .....	82, 89, 130, 164	Chen, Yan .....	81, 174
Chan, Shueng-Han Gary .....	85	Chen, Yi-Hau .....	52
Chan, Wei-Kai .....	168	Chen, Yi-Ping Phoebe .....	126
Chan, Yuk-Hee .....	58	Chen, Yih-Farn .....	123
Chang, Bin .....	109	Chen, Yu-Han .....	47
Chang, Cheng-Yu .....	56	Chen, Yu-Ting .....	77
Chang, Chuan-Wang .....	159	Chen, Yuanhao .....	93
Chang, Chun-Fa .....	170	Chen, Yung-Chang .....	158
Chang, Chun-Hao .....	99	Cheng, Chao-Chung .....	43
Chang, David Chih-Wei .....	47	Cheng, Chih-Chi .....	52
Chang, Hsui-Cheng .....	52	Cheng, Corey .....	75
Chang, Jing-Ying .....	43, 168	Cheng, En .....	145
Chang, Shih-Fu .....	48, 65, 86	Cheng, Irene .....	110
Chang, Yi .....	95	Cheng, Kan-Leung .....	82, 89
Chang, Yu-Lin .....	166	Cheng, Der-Chun .....	168
Chau, Lap-Pui .....	139, 146	Cheung, Gene .....	103, 158
Chau, Paul M. ....	44	Cheung, Sen-ching .....	139
Chaudhari, Upendra .....	93	Chevallet, Jean-Pierre .....	150, 163
Chávez, Edgar .....	143	Chhetri, Amit .....	88
Chbeir, Richard .....	54	Chi, HuiCheng .....	66
Chebil, Fehmi .....	115	Chi, Ming-Chieh .....	162
Chen, Chang Wen .....	65, 70, 161	Chia, Liang-Tien .....	112
Chen, Cheng-Yao .....	47, 51	Chiang, Chen-Kuo .....	38
Chen, Chih-Ming .....	158	Chiang, Wei-Kuo .....	106
Chen, Ching-Lun .....	153	Chiang, Yao-Yi .....	36

Chien, Lee-Feng .....	97	Di, Nan .....	30
Chien, Shao-Yi .....	43, 166, 168	Diepold, Klaus .....	44
Chitturi, Suresh .....	162	Ding, Wenpeng .....	81
Chiu, Chih-Yi .....	97	Dionisio, Carlos Ramon Pantaleon .....	143
Chiu, Chui-Wen .....	58	Divakaran, Ajay .....	98, 135
Chiu, Hsu-Kuang .....	166	Doi, Munehiro .....	57
Chiu, Yung-Feng .....	170	Dong, Aijuan .....	148
Chng, Engsiong .....	86, 153	Dong, Lan .....	77
Cho, Nam Ik .....	37	Dong, Ming .....	35
Cho, Sanghyup .....	142	Dowlatsahi, Mehran .....	109
Cho, Sukhee .....	173	Drew, Mark S. ....	45, 127
Choi, Hongsik .....	110	Driessen, Peter .....	144
Choi, Jean .....	131	Du, Jianguo .....	59
Choi, Minseok .....	113	Duan, Lingyu .....	61, 97
Chou, Shu-Hsuan .....	30	Dugelay, Jean-Luc .....	102, 125
Choudhury, Anustup .....	79	Duhamel, Pierre .....	101, 119, 124
Christodoulou, Lakis .....	140	Dulong, Carole .....	30, 63
Chu, Jui-Chin .....	30	Dumitrescu, Sorina .....	68, 155
Chu, Selina .....	87	Durkovic, Marko .....	44
Chu, Stephen .....	165	Ebadollahi, Shahram .....	86
Chu, Wei-Ta .....	148	Ebrahimi-Moghadam, Abbas .....	76
Chua, Tat-Seng .....	168	Eisert, Peter .....	34, 173
Chuah, Chen-Nee .....	158	El'arbi, Maher .....	132
Chuang, Shang-Chih .....	39	Elder, Bruce .....	97
Chung, Kelvin .....	36	Elgammal, Ahmed .....	34
Chung, Pau-Choo .....	56	El Saddik, Abdulmotaleb .....	174
Chung, Yun-Su .....	131	Engel, Dominik .....	174
Ci, Wang .....	125	Englert, Roman .....	34
Civanlar, M. Reha .....	173	Enzmann, Mitch .....	170
Coelho, Renata .....	98	Eom, Ki-Wan .....	100
Cohen, Michael .....	104	Erol, Berna .....	171
Colombo, Alessandro .....	40, 130	Erzin, Engin .....	87
Connell, Jonathan .....	31	Eslami, Ramin .....	59, 119
Constant, Patrick .....	146	Essannouni, Fedwa .....	44
Coumou, David J. ....	84	Eum, Nakwoong .....	113
Cuenca, Pedro .....	49	Ewerth, Ralph .....	144
Culibrk, Dubravko .....	171	Falquet, Gilles .....	150
Cusano, Claudio .....	40, 130	Fan, Xiaoming .....	154
Dai, Jufeng .....	102	Fan, Xin .....	54, 72
Dai, Li-Rong .....	91, 127	Fang, Xiangzhong .....	76
Dai, Qionghai .....	51, 67, 109, 174	Farrugia, Jean-Philippe .....	67
Das, Pankaj .....	78	Fatemi, Omid .....	163
Davis, Stephen .....	128	Fattahi, Ahmad Reza .....	86
Deller, John .....	119	Favetta, Franck .....	54
Delmas, Patrice .....	165	Fehn, Christoph .....	152, 173
Del Bimbo, Alberto .....	35, 50, 64	Feng, Huang .....	122
Dera, Thomas .....	170	Fernandez-Escribano, Gerardo .....	49
De Martin, Juan Carlos .....	121	Ferzli, Rony .....	62
De Mauro, Andrea .....	90	Fieguth, Paul .....	155
de With, Peter H.N. ....	152	Finlayson, Graham .....	45
Dhavale, Nitin .....	63	Florêncio, Dinei .....	88, 172



Fok, Apple W. P.	167	Guo, Xingfeng	109
Fotouhi, Farshad	35	Guo, Xun	40
Foures, Thomas	169	H. Mahoor, Mohammad	94
Freisleben, Bernd	144	H. N. de With, Peter	68
Frossard, Pascal	121, 128, 129, 144, 149	Hajdu, András	96
Fu, Fangwen	86	Hall, Varrian	107
Fu, Jingjing	75	Haller, Martin	148
Fujii, Toshiaki	36, 57	Hampapur, Arun	113
Fujiyoshi, Hironobu	71	Hamzaoui, Raouf	42
Fung, Yik-Hing	58	Han, Bo	69
Furht, Borko	133, 140, 171	Han, Chia	94
Gabbouj, Moncef	68	Han, Youn-Hee	69
Gagnon, Langis	46	Hannuksela, Miska M.	68
Galan, Jensen	51	Hare, Jonathon	63
Gamba, Humberto	78	Hasegawa, Yuichi	56
Gao, Li	145	Hashimoto, Hideo	76
Gao, Sheng	127	Hashimoto, Sadatsugu	153
Gao, Wen	33, 37, 38, 40, 52, 56, 59, 76, 83, 112, 125, 135	Hashiyama, Tomonori	71
Garcia, Christophe	47	Hata, Toshihiko	124
Gatica-Perez, Daniel	163	Haubold, Alexander	145
Georganas, Nicolas	43	Hauptmann, Alexander	32, 48, 160
Geusebroek, Jan-Mark	126	Hawellek, Michael	60
Ghaderpanah, Mohammadreza	132	He, Lei	94
Gharai, Ladan	57, 129	He, Li-Wei	104
Ghasemi, Hamid Reza	163	He, Li-wei	172
Ghoraani, Behnaz	132	He, Yifeng	128
Ghouti, Lahouari	159	He, Yuwen	37, 81, 125
Gibbon, David	123	Hellwagner, Hermann	39
Gimel'farb, Georgy	165	Ho, Chi-Wang	81, 84, 164
Girod, Bernd	66, 85	Ho, Chin-Chuan	159
Glasberg, Ronald	126	Ho, Yo-Sung	45, 116
Gleicher, Michael	126	Hoashi, Keiichiro	96, 144, 149
Gong, Yihong	154	Hongyi, Lu	53
Gong, Yuchang	72	Hopf, Klaus	140
Gordy, James	157	Horain, Patrick	67
Gosselin, Philippe-Henri	146	Horie, Arata	71
Goubran, Rafik	157	Hörster, Eva	112
Grangetto, Marco	118, 146	Hou, Jun	76
Gray, Robert M.	46	Hourunranta, Ari	115
Gritai, Alexei	97	Hsiao, Chih-Chieh	39
Gu, Irene Y.H.	120	Hsieh, Chaur-Heh	148
Guan, Ling	31, 97, 128	Hsieh, Ming-Shing	133
Guan, Xinping	103	Hsieh, Shih-Sen	41
Guang, Terence Hung Gih	35	Hsieh, Tien-Wei	99
Guehenneux, Erwan	67	Hsu, Chih-Lung	60
Guest, Clark C.	78	Hsu, Yu-Feng	65
Guimaraes, Silvio	98	Hu, Wei	135
Gündüz, Sule	79	Hu, Yiqun	112
Gunnarsson, Anders	120	Hu, Yu	157
Guo, Jiun-In	30, 52, 99	Hu, Yu Hen	143
Guo, Liwei	83	Hua, Guogang	161

Hua, Kien.....	145	Ji, Xiangyang.....	38
Hua, Xian-Sheng.....	91, 127, 144	Jia, Weijia.....	69, 167
Huang, Beilei.....	39	Jian, Guang-Tung.....	162
Huang, Cheng.....	122	Jiang, Gangyi.....	42
Huang, Chun-Rong.....	97	Jiang, Hao.....	127
Huang, Chung-Chieh.....	42	Jiang, Hongbing.....	140
Huang, Chung-Hsien.....	41	Jiang, Jintao.....	106
Huang, Chung-Lin.....	153	Jiang, ShuQiang.....	134
Huang, Feng.....	37	Jiang, Shuqiang.....	134
Huang, Hao.....	99	Jiang, Xiaoyi.....	152, 160
Huang, Jianjun.....	148	Jiau, Hewijin Christine.....	159
Huang, Kai-Chen.....	36	Jin, Hai.....	145
Huang, Pon-Jay.....	58	Jin, Hongxia.....	116
Huang, Qian.....	59, 83	Jin, Jesse S.....	61, 97
Huang, Qingming.....	33, 83, 134, 135	Jin, Xing.....	82, 89, 130
Huang, Szu-Hao.....	146	Jing, Feng.....	145
Huang, Thomas S.....	65, 95, 136, 154, 165	Joly, Philippe.....	169
Huang, Wendong.....	88	Joshi, Siddharth.....	171
Huerta, Juan.....	93	Jurca, Dan.....	121, 149
Hughes, Esther.....	110	Kakusho, Koh.....	43
Huh, Young.....	118	Kalman, Mark.....	172
Hull, Jonathan J.....	171	Kalva, Hari.....	33, 49, 140, 171
Hung, Hao-Ping.....	104	Kamath, Rohit.....	139
Hung, Mike.....	55	Kameyama, Seiji.....	169
Hung, Yi-Ping.....	55, 77	Kamikura, Kazuto.....	36
Hur, Namho.....	173	Kang, Chih-Heng.....	30
Ibarrola, Antonio C.....	143	Kankanhalli, Mohan.S.....	88
Ibrahim, Mohammad.....	159	Kankanhalli, Mohan S.....	77, 92, 147
Ide, Ichiro.....	126	Kao, Chao-Yang.....	111
Iiyama, Masaaki.....	43	Karam, Lina J.....	62, 121
Im, Sio Kei.....	124	Karande, Shirish.....	82, 158
Imamura, Kousuke.....	76	Karpov, Alexey.....	87
Ingrassia, Michael.....	162	Käster, Thomas.....	54
Ip, Horace H. S.....	49, 167	Katayama, Norio.....	126
Iqbal, Razib.....	174	Katsaggelos, Aggelos K.....	145
Ishibashi, Yutaka.....	169	Kauff, Peter.....	34, 152, 173
Islam, Asad.....	115	Kaup, André.....	32
Itti, Laurent.....	63	Kawada, Ryoichi.....	78
Iwata, Mitsuru.....	71	Kawamura, Kei.....	39
J. Sciabassi, Robert.....	160	Keim, Daniel.....	90
Jagmohan, Ashish.....	31	Kender, John R.....	31
Jamin, Sugih.....	122	Kerhervé, Brigitte.....	156
Jang, Jong Whan.....	46	Kesavarapu, Sreeni.....	31
Jarrah, Kambiz.....	31	Khaleghi, Azadeh.....	53
Jayant, Nikil.....	62	Khan, Saad.....	45
Jen, Chien-Wei.....	47	Khokhar, Ashfaq.....	51
Jeon, Hyun-Ho.....	144	Khoshgozaran-Haghighi, Jeff.....	36
Jeong, Hong.....	79	Ki, Myungseok.....	42
Jeong, Jechang.....	142	Kieffer, Michel.....	124
Jeong, Seyoon.....	42	Kim, Donghyung.....	142
Jeong, Yo-Won.....	124	Kim, Duckwhan.....	113

Kim, Eun Jung .....	107	Kumar, Vinay .....	147
Kim, Hae Yong.....	143	Kunath, Peter .....	120
Kim, HeeJeong .....	116	Kundur, Deepa.....	65
Kim, Hyoung-Gook .....	100, 148	Kuo, C.-C. Jay .....	72, 79, 87, 92, 110, 157
Kim, Il Koo.....	37	Kuo, Huang-Chih .....	111
Kim, Jae-kyoon.....	124	Kuok, Wai-He.....	105
Kim, Jaeseok.....	74	Kurozumi, Takayuki.....	51
Kim, Jinkyu .....	113	Kuwahara, Naoki .....	124
Kim, Jinwoong .....	173	Kwon, Do-Kyoung .....	72
Kim, Jin Ryong.....	67, 69	Kwon, Gisik.....	66
Kim, Ki-Hyun.....	131	Kwon, Jae Cheol.....	124
Kim, Kyoung Ill.....	107	Kwon, Ki-Hyeon .....	115
Kim, Kyuheon .....	42	Kwon, Ki-Ryong .....	118
Kim, Namyong .....	115	Kwong, Sam .....	80
Kim, Sang-Jun .....	115	Kyan, Matthew .....	31
Kim, Seongmin.....	113	Lahanier, Christian .....	146
Kim, Seonki.....	37	Lai, Edmund M-K .....	39
Kim, Shin-Hyoung .....	46	Lai, Shang-Hong.....	38, 146
Kim, Sung-Yeol.....	45	Lambers, Martin .....	152
Kimata, Hideaki.....	36	Landon, Paul.....	156
Kirenko, Ihor .....	59	Laurini, Robert .....	54
Kitahara, Masaki.....	36	Le, Lam.....	132
Klein Gunnewiek, Rene.....	68	Lee, Bong-Ho .....	116
Knoblock, Craig .....	36	Lee, Bu-Sung .....	112
Koba, Yoshimasa.....	76	Lee, Chan-Su .....	34
Koedinger, Kenneth R. ....	166	Lee, Changhyun.....	74
Koelma, Dennis .....	126	Lee, Cheng-Hsien .....	60
Koike, Atsushi.....	78	Lee, Gwo Giun .....	76
Kok, Chi-Wah.....	80	Lee, Heung Ki .....	107
Kokiopoulou, Effrosyni .....	144	Lee, Hsiao-Ping .....	105
Kolazhi, Krishnan.....	139	Lee, Injae .....	42
Kolekar, Maheshkumar .....	135	Lee, Ivan .....	128, 154
Kollias, Stefanos.....	117	Lee, Jae-Beom .....	33
Komiya, Ryoichi.....	56	Lee, Jenq-Kuen.....	47
Koo, Bontae.....	113	Lee, Jiann-Der .....	41
Koompairojn, Soontharee.....	145	Lee, Juhyun.....	113
Koskela, Markus.....	32	Lee, Moon-Chuen.....	61, 147, 155
Kotropoulos, Constantine .....	92, 101, 168	Lee, Moon-Hyun .....	115
Kotti, Margarita .....	101, 168	Lee, Sang-Kwang .....	101, 116
Koval, Oleksiy .....	83	Lee, Sekho .....	113
Kozintsev, Igor .....	63	Lee, Seongjoo .....	74
Kriegel, Hans-Peter .....	120	Lee, Seungjae .....	101
Krishnamurthy, Ganesh.....	164	Lee, Shin-Tseng.....	41
Krishnan, Sri.....	31	Lee, Soo-In .....	116, 173
Krishnan, Sridhar.....	132	Lee, Suh-Yin.....	159
Ku, Wei-Chun.....	30	Lee, Suk-Hwan .....	118
Ku, William .....	77	Lee, Sung-Ju .....	103
Kui, Dai .....	53	Lee, Sung-Sun .....	33
Kulkarni, Santosh .....	139	Lee, Sung Chun .....	36
Kumano, Masahito .....	114	Lee, Wan-Chi .....	92
Kumar, Anurag .....	147	Lee, Wing .....	140

Lee, Wonjun .....	59	Liang, Zhiqin .....	83, 174
Lee, Yung-Ki .....	33	Liao, I-Tao .....	47
Lee, Yung-Lyul .....	33	Liao, Shou-Chun .....	55
Leem, Lark-hoon .....	31	Liebl, Günther .....	69
Lehman, Tom .....	57	Lienhart, Rainer .....	112
Lei, Bao .....	68	Lifeng, Sun .....	122
Lelescu, Dan .....	80	Lim, Bum Hyun .....	67
Leonardi, Riccardo .....	93	Lim, Joo-Hwee .....	77, 150, 163
Leow, Wee Kheng .....	58	Lim, Seong-Jae .....	37
Leung, Henry .....	113	Lin, Bo-Yun .....	76
Leung, Man-Fung .....	85	Lin, Chang-Hong .....	123
Lewis, Paul .....	63	Lin, Chia-Wen .....	158
Li, Bin .....	37, 109	Lin, Chien-Chang .....	52, 99
Li, Bo .....	103	Lin, David Chih-Che .....	44
Li, Cheng-Hung .....	97	Lin, He-Yuan .....	76
Li, Chunwen .....	103	Lin, Ji-Kun .....	74
Li, Danjue .....	158	Lin, Lidong .....	69
Li, Eric .....	30	Lin, Louis H .....	59
Li, Fei .....	51	Lin, Po-Chiang .....	107
Li, Haizhou .....	92	Lin, Shouxun .....	50
Li, Hao .....	58, 102	Lin, Tsung-Nan .....	60, 107, 111
Li, Honglin .....	148	Lin, Wei-Hao .....	32, 160
Li, Houqiang .....	73, 161	Lin, Wei-Yang .....	143
Li, Hua-Fu .....	159	Lin, Weisi .....	73, 125
Li, Hui-Ya .....	106	Lin, Youn-Long .....	111
Li, Jiaming .....	112	Lindroth, Tuomas .....	112
Li, Jiang .....	110	Ling, Hefei .....	132, 134
Li, Jianguo .....	135	Liu, Chunxi .....	134
Li, Jiliang .....	76	Liu, David .....	35
Li, Jin .....	66, 122	Liu, Debing .....	37
Li, JinTao .....	73	Liu, Dong .....	81
Li, Mingjing .....	52, 54, 93, 145	Liu, Feng .....	126
Li, Qing .....	157	Liu, Hang .....	102
Li, Shan .....	147	Liu, Jia .....	137
Li, Shipeng .....	37, 73, 125, 144, 151, 161	Liu, Jiangchuan .....	121
Li, Songnan .....	59	Liu, Jing .....	61, 97
Li, Te .....	105	Liu, Jingen .....	93
Li, Wanqing .....	132	Liu, Kang .....	44
Li, Weihong .....	115, 137	Liu, Li-Chang .....	41
Li, Wenlong .....	30	Liu, Lurng-Kuo .....	31
Li, Xiang .....	93	Liu, Ming .....	65
Li, Xiaorong .....	35	Liu, Peiran .....	43
Li, Yan .....	85	Liu, Qiang .....	160
Li, Yang .....	73, 114, 129	Liu, Qingshan .....	38, 61
Li, Ze-Nian .....	127	Liu, Shao-Wei .....	38
Li, Zhenglong .....	97	Liu, Xiaodong .....	109
Li, Zhi .....	65	Liu, Xiaoqing .....	143
Li, Zhiwei .....	52, 93	Liu, Xin .....	67
Li, Zhu .....	145	Liu, Yan .....	112
Lian, Meifeng .....	97	Liu, Yebin .....	174
Lian, Yong .....	65	Liu, Yi-Chi .....	131

Liu, Ying.....	90	Mao, Mark .....	172
Liu, Zhengkai .....	73	Marcotte, Odile .....	156
Liu, Zhengye.....	81, 102	Marpe, Detlev .....	155
Liu, Zhu .....	65, 123	Marques, Oge .....	78, 140, 171
Liu, Zicheng .....	105	Marras, Ioannis .....	72
Lo, Chia-Jung .....	39	Martins, Luís Gustavo P.M. ....	101
Lo, Hung-Yi .....	60	Martins, Pedro .....	133
Logerot, Guillaume.....	146	Mase, Kenji.....	57
Loke, Mei Hwan .....	73	Mataric, Maja .....	87
Long, Chengnian .....	103	Matsumoto, Kazunori.....	96, 144, 149
Long, Darrell .....	111	Matsuoka, Akihiro.....	133
Long, Rodney .....	160	Matta, Federico.....	125
Loscoc, Alex .....	167	Mavandadi, Sam .....	53
Lotspiech, Jeffery .....	116	Mayilavelane, Aroutchelvame.....	113
Lou, Jian-Guang .....	110	Mayron, Liam .....	78, 140
Lu, Cheng .....	45	McKittrick, Chad .....	115
Lu, Haibin.....	140	Mega, Satoru.....	71
Lu, Haiping.....	95	Mei, Tao .....	144
Lu, Hanqing.....	38, 61, 86, 97	Merabti, Madjid.....	65
Lu, Juwei .....	108	Merialdo, Bernard.....	94
Lu, Mei-Hsuan.....	170	Merkle, Philipp .....	142, 173
Lu, Shi .....	89	Metaxas, Dimitris .....	34
Lu, Wenming.....	132	Migliorati, Pierangelo.....	93
Lu, Yan .....	138	Mikami, Dan.....	53
Lu, Zhengding .....	132, 134	Ming, Anlong .....	131
Lu, Zhongkang .....	73	Mingche, Lai.....	53
Lukac, Rastislav .....	82, 107	Minoh, Michihiko.....	43
Luo, Jian-Guang .....	89, 109	Misra, Kiran.....	158
Luo, Jiebo .....	79	Miyamori, Hisashi .....	54
Lyrroudia, Kleoniki .....	72	Mo, Hiroshi .....	126
Lyu, Michael.....	89	Mohamed, Magdi .....	45
Ma, Huadong .....	131	Mohd Arif, Nor Azhar .....	56
Ma, Mengyao.....	83	Mojon, Daniel.....	160
Ma, Siwei.....	38, 157	Mooser, Jonathan.....	115
Ma, Songde.....	61	Mori, Kensaku .....	57
Ma, Wei-Ying.....	52, 72, 93, 145	Mori, Shinya .....	100
Ma, Xiang .....	51	Mori, Yumi .....	57
Ma, Xin .....	83	Morimoto, Masashi.....	53
Macaire, Ludovic.....	94	Morita, Satoru .....	41
Machida, Takashi .....	71	Moriyama, Tsuyoshi .....	100
Maddage, Namunu C.....	92	Mueller, Karsten .....	173
Magli, Enrico.....	118	Muharemagic, Edin .....	133
Mahdi, Walid.....	143	Muijs, Remco .....	59
Mahmoud, Qusay .....	72	Müller, Karsten .....	142
Malkin, Robert.....	171	Müller, Marcus .....	152
Malvar, Rico .....	104	Muntean, Gabriel-Miro.....	69
Mamalet, Franck.....	47	Murase, Hiroshi .....	92
Man, Hong .....	129	Murdoch, Jennifer.....	91
Manders, Corey .....	108	Muselet, Damien.....	94
Manjikian, Naraig.....	164	N. Kachouie, Nezamoddin .....	155
Mann, Steve.....	55, 108	Nagata, Hidenobu .....	53

Nahrstedt, Klara.....	70	Oulad Haj Thami, Rachid.....	44
Naik, Sagar.....	81	Ou Yang, Yi-Huan.....	99
Naito, Masaki.....	144, 149	Owada, Shigeru.....	56
Nakajima, Yasuyuki.....	78, 117, 133	Ozawa, Shinji.....	100, 114, 153
Nakamura, Naoto.....	42	Ozbek, Nukhet.....	43
Nakamura, Yuichi.....	50	Ozeki, Motoyuki.....	50
Nakano, Hiroki.....	57	Özsu, Tamer.....	79
Nanri, Yasuhito.....	41	Pai, Yu-Ting.....	131
Naphade, Milind R. ....	31, 145	Pala, Pietro.....	35
Narayanan, Shrikanth.....	87	Pan, Pan.....	59
Nasiopoulos, Panos.....	74	Pan, Yunhe.....	166
Natsev, Apostol (Paul).....	78, 145	Panse, Christian.....	90
Navarro, Antonio.....	161	Pantic, Maja.....	50
Nesi, Paolo.....	57, 67, 118	Panwar, Shivendra.....	81
Neumann, Ulrich.....	36, 115	Papaleontiou, Leontios.....	72
Nevatia, Ram.....	36	Pâris, Jehan-François.....	111, 139
Ng, Teck Khim.....	58	Park, Hanhoon.....	115
Ngo, Chong-Wah.....	71	Park, Jong-Il.....	115
Nguyen, Thinh.....	139	Park, Junghun.....	110
Ni, Zefeng.....	138	Park, Kihyuk.....	113
Nicolas, Henri.....	132	Park, Kyu-Ho.....	124
Nielsen, Frank.....	56	Park, So Ra.....	116
Nijjima, Koichi.....	42	Parrikar, Utpal.....	158
Nikolaidis, Nikolaos.....	72	Paulovicks, Brent.....	31
Nikolaidis, Nikos.....	147	Pearmain, A. J.....	124
Nikolopoulos, Spyros.....	147	Pehlivan, Selen.....	173
Ning, Huazhong.....	154	Peker, Kadir.....	98
Nishida, Shin'ya.....	62	Peng, Chengyuan.....	70
Nishiyama, Hiroki.....	57	Peng, Lina.....	66
Niskanen, Matti.....	55	Penn, Gerald.....	80
Nitta, Naoko.....	87	Perkins, Colin.....	57, 129
Noh, Jeonghun.....	66	Persaud, Krishna.....	115
Ntalianis, Klimis.....	117	Petracca, Matteo.....	121
Nunziati, Walter.....	64	Petrovic, Goran.....	152
Obermeier, Florian.....	44	Pfeiffer, Michael.....	54
Ofli, Ferda.....	87	Philipp-Foliguet, Sylvie.....	146
Ogunbona, Philip.....	132	Phua, Koksoon.....	156
Oh, Yunje.....	74	Pietquin, Olivier.....	56
Ojala, Timo.....	53	Pighin, Frédéric.....	63
Okada, Ken-ichi.....	114	Pilati, Laurent.....	41
Okada, Yoshihiro.....	42	Ping, Xijian.....	119
Ollikainen, Ville.....	70	Ping, Xue.....	125
Olmo, Gabriella.....	118, 146	Pinhas, Adi.....	30
Ong, Ee Ping.....	73	Pinzón, Juan.....	51
Orozco-Barbosa, Luis.....	49	Pitas, Ioannis.....	72, 95, 96, 147
Ortimini, L.....	67	Plataniotis, Konstantinos N.....	82, 95, 107, 108, 136
Ostermann, Joern.....	44	Platt, John.....	120
Otsuka, Isao.....	98, 135	Pong, Ting-Chuen.....	71
Otsuka, Kazuhiro.....	92	Poon, Eunice.....	108
Ou, Jiazhi.....	104	Potamianos, Gerasimos.....	165
Ouali, Anis.....	156	Poucet, Christophe.....	98

Prangl, Martin.....	39	Sapino, Maria Luisa.....	66
Pryakhin, Alexey.....	120	Sarchi, Mauro.....	151
Pun, Thierry.....	83	Sardashti, Somayeh.....	163
Qi, Guo-Jun.....	91	Sargin, Mehmet Emre.....	87
Qi, Honggang.....	38	Sat, Batu.....	172
Qi, Yingyi.....	58	Satoh, Shin'ichi.....	29, 126
Qiu, Bo.....	97	Sattar, Farook.....	98
Qu, Wei.....	45, 95	Saurin, Alvaro.....	57
Raahemifar, Kaamran.....	113	Schettini, Raimondo.....	40, 130
Radha, Hayder.....	59, 82, 119, 158	Schierl, Thomas.....	69
Radhakrishnan, Regunathan.....	135	Schlessman, Jason.....	47, 123
Radhouani, Saïd.....	150	Schneider, Erich.....	170
Rafi, Fahd.....	45	Schonfeld, Dan.....	45, 51, 59, 90, 95
Rahardja, Susanto.....	29, 105	Schreck, Tobias.....	90
Rahman, A. M. Tushar Iqbal.....	164	Schreer, Oliver.....	34, 152
Rajan, Deepu.....	112	Schreier, Ralf M.....	164
Rama, Antonio.....	136	Schuller, Bjoern.....	29, 46
Ramakrishnan, K.R.....	116	Schuller, Björn.....	60, 92, 100
Ramaswamy, Ganesh.....	93	Schwartz, Stuart C.....	77
Ramliy, Mohd Nazri.....	56	Schwarz, Heiko.....	155
Randall, Jonathan.....	154	Schwenke, Derek.....	124
Rao, Sira.....	62	Sebe, Nicu.....	64
Rautiainen, Mika.....	53	Secoleanu, Tiberiu.....	112
Rege, Manjeet.....	35	Seinstra, Frank.....	126
Reisslein, Martin.....	121	Sekiguchi, Shun-ichi.....	49
Reiter, Stephan.....	29, 46, 92	Sengupta, Somnath.....	135
Rezaei, Mehdi.....	68	Seo, Young-Ho.....	101
Rigoll, Gerhard.....	29, 46, 60, 92, 100	Seppänen, Tapio.....	53
Rihan, Arjun.....	36	Sert, Mustafa.....	91
Ritz, Christian.....	128	Servetti, Antonio.....	121
Röder, Martin.....	155	Setlur, Vidya.....	162
Rogai, Davide.....	67, 118	Setton, Eric.....	66
Ross, Keith.....	81	Sexton, Ian.....	140
Rossello, Norbert.....	41	Shah, Mubarak.....	45, 93, 97
Roth, Gerhard.....	43	Shahabi, Cyrus.....	36
Rothermel, Albrecht.....	164	Shahraray, Behzad.....	123
Roubies, Athanasios.....	96	Shan, Man-Kwan.....	159
Roux, Sébastien.....	47	Shao, Feng.....	42
Roy, Deb.....	163	Shao, Ling.....	59
Ru, Congchong.....	70	Shao, Xi.....	88
Ruan, Shanq-Jang.....	131	Sharifi, Ahmedullah.....	55
Rudolph, Eric.....	104	Sharma, Gaurav.....	84
Rui, Yong.....	70, 104, 136, 172	Sharma, Puneet.....	103
Ryan, Ken.....	46	Sheinin, Vadim.....	31
S., Prasad.....	116	Shen, Bo.....	49, 172
Safaei, Farzad.....	109	Shen, Day-Fann.....	58
Safavi-Naini, Rei.....	132	Shen, Guobin.....	125
Sakauchi, Masao.....	29	Shen, Huifeng.....	161
Salam, Ahmed.....	44	Shen, Mei-Yin.....	72
Samarabandu, Jagath.....	143	Shen, Xuemin.....	81
Sanigepalli, Praveen.....	140	Shen, Yanming.....	81



Shi, Yuan-Chun .....	109, 167	Sugimoto, Kazuo .....	49
Shi, Yuan-Yuan .....	100	Suh, Young-Ho .....	116
Shi, Yun .....	118	Sun, Bing-Yu .....	61
Shie, Mon-Chau .....	131	Sun, Lifeng .....	37, 130
Shigeno, Hiroshi .....	114	Sun, Ming-Ting .....	141
Shih, Huang-Chia .....	153	Sun, Mingliang .....	135
Shih, Timothy K. ....	59	Sun, Mingui .....	160
Shih, Zen-Chung .....	55	Sun, Qibin .....	29, 64, 65, 127
Shim, Kwang Hyun .....	67, 69	Sun, Tiantian .....	76
Shimizu, Shinya .....	36	Sun, Wen .....	138
Shiqiang, Yang .....	37, 122	Sun, Xiaoyan .....	73, 161
Shirani, Shahram .....	76, 77	Sun, Yanfeng .....	74
Shirmohammadi, Shervin .....	174	Surman, Phil .....	140
Shiu, Yu .....	79	Szkaliczki, Tibor .....	39
Shoa, Alireza .....	77	Tagliasacchi, Marco .....	151
Shue, Louis .....	156	Takayama, Makoto .....	117
Sidiropoulos, Panagiotis .....	147	Takayama, Tatsuhiko .....	71
Sikora, Thomas .....	126, 148	Takeda, Kazuya .....	57
Silva, Antonio .....	161	Takemae, Yoshinao .....	92, 114
Silvén, Olli .....	55	Takemura, Haruo .....	71
Singh, Ajit .....	81	Takimoto, Masao .....	29
Singh, Jaswinder Pal .....	47	Tam, Wa James .....	152
Singh, Rahul .....	51	Tan, Wai-tian .....	158, 172
Sinnathamby, Mohanarajah .....	164	Tanaka, Katsumi .....	54
Smeaton, Alan .....	32	Tanaka, Kiyoshi .....	117, 133
Smeulders, Arnold .....	126	Tang, Hao .....	95, 115
Smith, Jason .....	151	Tang, Jiayu .....	63
Smith, John R. ....	86, 127	Tang, Lijung .....	31
Smolic, Aljoscha .....	142, 152, 173	Tang, Xiaouu .....	97
Snoek, Cees .....	126	Tang, Yong .....	110
Socek, Daniel .....	171	Tang, Yun .....	89, 130
Sohn, Jae-Sik .....	118	Tanger, Ralf .....	34, 152
Solachidis, Vassilios .....	95	Tanimoto, Hiroaki .....	87
Song, Haohao .....	58	Tanimoto, Masayuki .....	36, 57
Song, Li .....	58, 73	Tano, Shun'ichi .....	71
Song, Mingwu .....	119	Tao, Dayong .....	150
Song, Yan .....	91, 127	Tao, Jianhua .....	137
Song, Ying .....	156	Tarres, Francesc .....	136
Sprague, Alan .....	90	Tas, Cengiz .....	126
Stockhammer, Thomas .....	69	Tashev, Ivan .....	104
Stokes, Jack .....	88, 120	Taube, Wolfgang .....	51
Strintzis, Michael G. ....	68	Tavares, Joao .....	161
Su, Ching-Lung .....	52	Tawada, Masaki .....	169
Su, Drew Wei-Chi .....	76	Tekalp, A. Murat .....	43, 87
Su, Guan-Ming .....	104	Tešić, Jelena .....	127
Su, Wei .....	118	Teuhola, Jukka .....	112
Su, Yih-Ming .....	148	Thakkar, Snehal .....	36
Subramanya, Amarnag .....	105	Thiran, Jean-Philippe .....	40
Sudharsanan, Subramania .....	164	Thoma, George .....	160
Suenaga, Yasuhito .....	57	Thomas-Kerr, Joseph .....	128
Sugaya, Fumiaki .....	96, 144, 149	Thomos, Nikolaos .....	68

Tian, Qi.....	48, 61, 64, 97	Wagner, Jean-Paul.....	128
Tian, Ying-Li.....	113	Wagner, Philipp.....	170
Tico, Marius.....	55	Wah, Benjamin W. ....	172
Tillo, Tammam.....	146	Waibel, Alex.....	171
Tjondronegoro, Dian.....	126	Wallhoff, Frank.....	46, 60, 100
Tong, Xiaofeng.....	86	Wan, Kongwah.....	153
Torres, Anne.....	98	Wang, Bin.....	52, 144
Toyoura, Masahiro.....	43	Wang, Chen.....	58
Trichet, Rémi.....	94	Wang, Chia-Wei.....	148
Tsai, Chuan-Yung.....	156	Wang, Chiapin.....	107, 111
Tsai, Shu-Min.....	41	Wang, Chou-Chen.....	39
Tsai, Yu-Pao.....	55	Wang, Chun-hao.....	97
Tseng, Din-Chang.....	42, 133	Wang, Chunli.....	56
Tseng, Shau-Yin.....	47, 99	Wang, Dan.....	121
Tu, Jilin.....	165	Wang, Feng.....	71
Tu, Yu-Kuang.....	141	Wang, Haifeng.....	38
Tubaro, Stefano.....	151	Wang, Hanli.....	80
Tung, Yi-Shin.....	39	Wang, Hung-Ming.....	74
Tyan, Hsiao-Rong.....	110	Wang, Jhing-Fa.....	56, 105
Typke, Rainer.....	146	Wang, Jia.....	38
Tzanetakis, George.....	91	Wang, Jia-Ching.....	105
Tzouveli, Paraskevi.....	117	Wang, Jie.....	136
Uddenfeldt, Mats.....	96	Wang, Jing-Wein.....	149
Uehara, Kuniaki.....	71, 114	Wang, Jinjun.....	86, 153
Uhl, Andreas.....	174	Wang, Jinn-Shyan.....	99
Ukai, Kazuhiko.....	141	Wang, Jinqiao.....	61, 97
Urfalioglu, Onay.....	44	Wang, Ju.....	110
Valli, Alessandro.....	50	Wang, Meng.....	127
Vallotti, Andrea.....	67, 118	Wang, Ming-Jiun.....	76
Valstar, Michel.....	50	Wang, Qi.....	56
Van der Auwera, Geert.....	121	Wang, Qing.....	36
van der Schaar, Mihaela.....	86	Wang, Ren-Hua.....	91
Vazhenin, Alexander.....	99	Wang, Sashuang.....	102
Vazhenin, Dmitry.....	99	Wang, Sung-Wen.....	39
Vedantham, Ramakrishna.....	162	Wang, Tao.....	30, 135
Veeravalli, Bharadwaj.....	35	Wang, Wen-Hao.....	146
Velipasalar, Senem.....	47, 113, 123	Wang, Wenjie.....	122
Veltkamp, Remco C.....	146	Wang, Xiaofeng.....	117
Venetsanopoulos, Anastasios (Tas) N.....	95, 136	Wang, Xiaohan.....	123
Venkatesh, Svetha.....	162	Wang, Yao.....	81, 102
Verbeeten, Russell.....	55	Wang, Ye.....	88, 167
Verscheure, Olivier.....	93	Wang, Yi.....	73
Vetro, Anthony.....	49, 124	Wang, Yongfang.....	69
Vilagut, Roger.....	40	Wang, Yongjin.....	97
Vinciarelli, Alessandro.....	147	Wang, Yu.....	139
Vitale, Ran.....	30	Wang, Zhe.....	68
Volkmer, Timo.....	78	Wang, Zhihang.....	33
Voloshynovskiy, Sviatoslav.....	83	Watanabe, Hiroshi.....	39
von dem Knesebeck, Matthias.....	74	Weda, Hans.....	134
Vretos, Nicholas.....	95	Wee, Susie.....	64, 172
Vu, Khanh.....	145	Wee, William.....	94

Wei, Bin.....	123	Xu, Jizheng.....	76, 151
Wei, Shou-Der.....	38	Xu, Min.....	112
Wei, Zhicheng.....	102	Xu, Wei.....	154
Weissenfeld, Axel.....	44	Xu, Wenli.....	51, 96, 174
Wen, Ying-Hsiang.....	156	Xu, Xun.....	136, 172
Wiegand, Thomas.....	69, 142, 155, 173	Xu, Yi.....	60
Wiering, Frans.....	146	Xuan, Guorong.....	118
Wilson, Stephen.....	87	Xue, Jianxia.....	106
Wolberg, George.....	137	Yamamoto, Kenji.....	36
Wolf, Wayne.....	47, 123	Yamamoto, Yuki.....	39
Wong, Alexander.....	75	Yamanoue, Hirokazu.....	141
Wong, Hau-San.....	96, 154, 165	Yamasaki, Toshihiko.....	29, 154
Wong, Hoi-Ming.....	84, 164	Yamato, Junji.....	51, 92
Wong, Kin-Chung.....	143	Yamazawa, Kazumasa.....	169
Wong, Wai-Choong.....	64	Yan, Rong.....	48, 95
Woodward, Alexander.....	165	Yang, Chung-Hsien.....	56, 105
Worring, Marcel.....	126	Yang, Heng.....	36
Wu, Bo.....	40	Yang, Jar-Ferr.....	41, 56, 74, 141
Wu, Chi-Jiunn.....	146	Yang, Jie.....	34, 95, 104, 166, 171
Wu, Chia-Hao.....	108	Yang, Jiheng.....	74
Wu, Chien-Tsai.....	41	Yang, Kai-Chieh.....	78
Wu, Di.....	52	Yang, Min.....	100
Wu, Fan.....	113	Yang, Shi-Qiang.....	89, 109
Wu, Feng.....	40, 73, 76, 81, 138, 151, 161	Yang, Shiqiang.....	130
Wu, Ja-Ling.....	39, 148	Yang, Weiyl.....	104
Wu, Jiangqin.....	166	Yang, Xiaokang.....	60, 69, 83
Wu, Kui.....	146	Yang, Yao-Chang.....	52
Wu, Min.....	104, 161	Yang, Yingchun.....	100
Wu, Qiufeng.....	67	Yang, Zhigang.....	112
Wu, Wen.....	34	Yanof, Arnold.....	112
Wu, Xiaolin.....	68, 123, 155	Yao, Jian.....	160
Wu, Yen-Nien.....	55	Yao, Susu.....	73
Wu, Yi.....	63	Yap, Kim-Hui.....	139, 146
Wu, Yingfei.....	166	Yashima, Yoshiyuki.....	36
Wu, Yongdong.....	117	Yasinnik, Yelena.....	87
Wu, Yung-Gi.....	108	Yazici, Adnan.....	91
Wu, Zhaohui.....	100	Yeh, You-Sheng.....	56
Wu, Zhimei.....	73	Yemez, Yucel.....	87
Wyse, Lonce.....	44	Yendo, Tomohiro.....	36
Xiao, Xin.....	167	Yeo, Hangu.....	31
Xie, Don.....	52	Yetongnon, Kokou.....	54
Xie, Lexing.....	48, 86	Yin, Baocai.....	74
Xie, Xing.....	72	Yin, Haibin.....	76
Xin, Jun.....	49	Yin, Lijun.....	151
Xiong, Hongkai.....	58, 73	Ying, Lihang.....	110
Xiong, Ruiqin.....	151	Yip, Shu-Kei.....	84, 164
Xiong, Yongqiang.....	122, 130	Yiu, W.-P. Ken.....	130
Xu, ChangSheng.....	97	Yokoi, Takao.....	71
Xu, Changsheng.....	86, 88, 135, 153	Yokoya, Naokazu.....	71, 169
Xu, Jianfeng.....	29	Yoneyama, Akio.....	117, 133
Xu, Jin.....	73	Yoo, Chang D.....	101

Yoo, Jang-Hee .....	131	Zhang, Xiao-Ping .....	117
Yoo, S. J. Ben .....	158	Zhang, Xiaodong .....	172
Yoon, Sangho .....	46, 172	Zhang, Xiaopeng .....	58
Yoon, Young-Suk .....	45	Zhang, Yan .....	119
You, Mingyu .....	137	Zhang, Yimin .....	30
You, Suyu .....	36, 115	Zhang, Yiming .....	135
You, Yilun .....	163	Zhang, YongDong .....	73
Yu, Bin .....	70	Zhang, Yongdong .....	50
Yu, Cheng-Wei .....	39	Zhang, Zhengyou .....	105
Yu, Feiyang .....	49	Zhang, Zhenqiu .....	165
Yu, Heather .....	65	Zhang, Zhishou .....	64
Yu, Jie .....	48, 64	Zhang, Zhongfei .....	160
Yu, Lun .....	70	Zhao, Debin .....	33, 37, 38, 40, 59, 83
Yu, Mei .....	42	Zhao, Zhao .....	134
Yu, Rongshan .....	105	Zheng, Haifeng .....	70
Yu, Songyu .....	58, 69, 73, 76, 83	Zheng, Yantao .....	61
Yu, Yanwei .....	132	Zhiying, Wang .....	53
Yu, ZhiWen .....	165	Zhong, Yuzhuo .....	130
Yu, Zhiwen .....	96, 154	Zhou, He-Qin .....	144
Yuan, Hua .....	117	Zhou, Jiantao .....	81, 174
Yuan, Wugang .....	134	Zhou, ShuMin .....	73
Yum, Ki Hwan .....	107	Zhou, Yinghua .....	72
Yun, Kugjin .....	173	Zhu, GuangYu .....	134
Zad-Issa, Mohammad .....	41	Zhu, Guangyu .....	135
Zafeiriou, Stefanos .....	147	Zhu, Jie .....	99
Zapata Ferrer, Maria .....	134	Zhu, Peng .....	103
Zeng, Bing .....	75, 150	Zhu, Xiaodan .....	80
Zeng, Hui-Chi .....	146	Zhu, Xiaoqing .....	85
Zeng, Wenjun .....	103, 140	Zhu, Xinglei .....	44
Zettl, Stefan .....	46	Zhu, Xuan .....	100
Zhai, Guangtao .....	60	Zhu, Yingnan .....	140
Zhai, Yun .....	93	Zhu, Yongwei .....	29
Zhang, Cha .....	70	Zhu, Zhigang .....	115, 137
Zhang, Chengcui .....	90	Zhuang, Xinhua .....	129
Zhang, Dong .....	163	Zhuang, Yueting .....	166
Zhang, Dongdong .....	83	Zoghلامي, Imad .....	77
Zhang, Enwei .....	96	Zou, Dekun .....	118
Zhang, Jiang .....	89, 109	Zou, Fuhao .....	132, 134
Zhang, Jing .....	34	Zuo, Li .....	110
Zhang, Jun .....	163	Zwick, Michael .....	44
Zhang, Li .....	125		
Zhang, Liang .....	138, 152		
Zhang, Meng .....	130		
Zhang, Nan .....	74, 109, 167		
Zhang, Peng .....	52		
Zhang, Qian .....	66, 122, 130		
Zhang, Tao .....	119		
Zhang, Wei .....	150		
Zhang, Weidong .....	96		
Zhang, Weigang .....	134		
Zhang, Wenjun .....	60, 83		