09w5056
Multimedia, Mathematics and Machine Learning II
5 July - 10 July, 2009

Organizers: Rabab Ward (UBC) and Li Deng (Microsoft)

*Breakfast (Buffet): 7:00 – 9:30 am, Sally Borden Building, Monday – Friday
*****Lunch (Buffet): 11:30 am – 1:30 pm, Sally Borden Building, Monday – Friday
*****Dinner (Buffet): 5:30 – 7:30 pm, Sally Borden Building, Sunday – Thursday
Coffee Breaks: As per daily schedule, 2nd floor lounge, Corbett Hall
****Please remember to scan your meal card in the dining room for each meal.***

MEETING ROOMS
All lectures will be held in Max Bell 159 (Max Bell Building accessible by walkway on 2nd floor of Corbett Hall). LCD projector, overhead projectors and blackboards are available for presentations.

Sunday
16:00 Check-in begins (Front Desk, Professional Development Centre, open 24 hours)
17:30-19:30 Buffet Dinner
20:00 Informal gathering in 2nd floor lounge, Corbett Hall
Beverages and small assortment of snacks available on a cash honour-system.

Monday
7:00-8:45 Breakfast

8:45-9:00 Introduction and Welcome to BIRS by BIRS Station Manager, Max Bell 159
9:00- 10:00 Bernd Girod
10:00- 10:30 Group Photo, meet on the front steps of Corbett Hall
10:30 -11:00 Coffee Break, 2nd floor lounge, Corbett Hall

11:00 -11:25 Lina Karam
11:25 -11:50 Jie Liang

11:50-13:00 Lunch (Buffet Lunch is available from 11:30 am to 1:30 pm daily)
13:00-14:00 Guided Tour of The Banff Centre; meet in the 2nd floor lounge, Corbett Hall

14:15-14:45 Jose Moura

2:45 – 3:15 Coffee Break
3:15- 3:40 Nick Kingsbury
3:40- 4:15 Ray Liu
4:15- 4: 40 Tsuhan Chen

Tuesday
7:00-9:00 Breakfast

9:00 –10:15 Hermann Ney ,Tutorial
10:15-10:45 Coffee Break,
10:45- 11:10 Li Deng
11:10 -11:35 Xiaodong He
11:35 -12:00 Dong Yu

12:00 - 1:30 Lunch

1:30- 1:55 Dick Lyon
1:55 -2:25 George Tzanetakis
2:25 - 2:50 Les Atlas

2:50-3.30 Coffee Break

17:30-19:30 Dinner

Wednesday
7:00-9:00 Breakfast

9:00 -9:25 Sabine Susstrunk
9:25- 9:50 Amir Said
9:50- 10:15 Ton Kalker

10.15-11.00 Coffee Break
Free the rest of the day

11:30-13:30 Lunch

17:30-19:30 Dinner

Thursday
7:00-9:00 Breakfast

9:00 - 10:00 Ed Delp

10:00 -10:30 Coffee Break
10:30 -11:30 Mehrdad Fatourechi, Jane Wang , Vicky Zhao
11:30 – 11:55 Darko Kirovski

11:55-13:30 Lunch

1:30-1:55 Deepa Kundur
1:55 - 2:20 Jeffrey Bloom
2:20- 2:40 Yucel Altunbasak
2:40-3.20 Coffee Break
3.20-3.45 Magdy Bayoumi

17:30-19:30 Dinner

Friday
7:00-9:00 Breakfast

Checkout by 12 noon.
** 5-day workshops are welcome to use the BIRS facilities (2nd Floor Lounge, Max Bell Meeting Rooms, Reading Room) until 3 pm on Friday, although participants are still required to checkout of the guest rooms by 12 noon. **

09w5056

Multimedia, Mathematics and Machine Learning II
5 July - 10 July, 2009

Organizers: Rabab Ward (UBC) and Li Deng (Microsoft)

ABSTRACTS

MONDAY

Bernd Girod
Lina Karam
Jie Liang
Jose' M F Moura
Nick Kingsbury
Tsuhan Chen
Ray Liu

Mobile Image Matching – Recognition Meets Compression

Bernd Girod
Stanford University

Handheld mobile devices, such as camera phones or PDAs, are expected to become ubiquitous platforms for visual search and mobile augmented reality applications. For mobile image matching, a visual data base is typically stored at a server in the network. Hence, for a visual comparison, information must be either uploaded from the mobile to the server, or downloaded from the server to the mobile. With relatively slow wireless links, the response time of the system critically depends on how much information must be transferred in both directions. We review recent advances in mobile matching, using a "bag-of-visual-words" approach with robust feature descriptors, and show that dramatic speed-ups are possible by considering recognition and compression jointly. We will use real-time implementations for different example applications, such as recognition of landmarks or CD cover, to show the benefit from image processing on the phone, the server, and/or both.

Adaptive Rate-Distortion Based Wyner-Ziv Video Coding

Lina Karam
Arizona State University

This talk starts with a brief introduction to the area of Distributed Video Coding (DVC), which is also known as Wyner-Ziv Video Coding. Two novel adaptive DVC systems are then presented: a pixel-domain DVC system with a rate-distortion based BitpLAne SelecTive decoding (BLAST-DVC), and a transform-domain DVC system with a rate-distortion based Adaptive QuanTization (AQT-DVC). Coding results and comparisons with existing DVC schemes and with H.264 interframe and intraframe coding are presented to illustrate the performance of the proposed systems.

Rectification-based View Interpolation and Extrapolation for Multiview Video Coding: R-D Analysis and Applications

Jie Liang
Simon Fraser University

View interpolation has been applied in multiview view coding (MVC). However, existing schemes assume all cameras are aligned. These methods do not perform well when neighboring cameras point to different directions. In this talk, we first derive the theoretical R-D performance of the rectification-based view interpolation. We then apply it to H.264 MVC. To further improve the coding efficiency, we develop a rectification-based view extrapolation for MVC. Finally, we investigate the application of the view interpolation in multiple-description coding of multiview images.

High Dimensional Consensus

José M. F. Moura
Carnegie Mellon University

This talk considers distributed algorithms that can arise when a large number of agents cooperate to reach a common decision or in sensor networks where a large number of sensors cooperate to process large amounts of collected data. In the last few years there has been intensive research in distributed algorithms for such problems. I will describe in this talk a general class of distributed algorithms, high dimensional consensus (HDC), that we have recently introduced. I will show how a number of problems of interest including distributed inference (like detection, estimation, or classification,) distributed localization, or several types of consensus algorithms can be cast in the framework of HDC. I will discuss the convergence of HDC under a broad set of conditions: deterministic as well as random, as when there is noise in the intersensor communications or links among sensors fail at random times. Finally, I will address tradeoffs among network and application parameters and their impact on resource allocation, convergence rate, and topology design.

Rotation-invariant wavelet-based matching of local features, with enhanced tolerance to shifts in location and scale.

Nick Kingsbury
University of Cambridge, UK

We describe a technique for using dual-tree complex wavelets to obtain rich feature descriptors of keypoints in images. The main aim has been to develop a method for retaining the full phase and amplitude information from the complex wavelet coefficients at each scale, while presenting the feature descriptors in a Fourier-domain form that allows for efficient correlation at arbitrary rotations between the candidate and reference image patches. The feature descriptors are known as Polar-Matching matrices. Recently, we have modified our previously proposed approach so that it can be more resilient to errors in keypoint location and scale. These multi-scale feature descriptors are potentially useful for object detection, recognition, classification and tracking in images and video

A Graphical-Model Framework for Using Context to Understand Images of People

Tsuhan Chen
Cornell University

When we see other humans, we can quickly make judgements regarding many aspects, including their demographic description and identity if they are familiar to us. We can answer questions related to the activities of, emotional states of, and relationships between people in an image. We draw conclusions based not just on what we see, but also from a lifetime of experience of living and interacting with other people. We propose contextual features and graphical models for understanding images of people with the objective of providing computers with access to the same contextual information that humans use.

Genomic/Proteomic Signal Processing for Cancer Classification and Prediction

Ray Liu
University of Maryland, College Park

DNA microarray and proteomic mass spectrum technologies make it possible to simultaneously monitor thousands of genes/protein expression levels and distribution. A topic of great interest is to study the different expression profiles from cancer patients and normal subjects, by classifying them at gene/protein expression levels. Currently, various clustering methods have been proposed in the literature to classify cancer and normal samples based on microarray data, and they are dominantly data-driven approaches. In this talk, an alternative model-driven approach, named ensemble dependence model, is presented aiming at exploring the group dependence relationship of gene clusters. Because of the limited size of current data, it is not feasible to examine

the regulation relationship between all genes. Also, both the microarray gene expression and mass spectrum data are noisy. However, if they are clustered in a right way, the noise level in the resulting cluster expression will be reduced, thus the ensemble dependence dynamics of gene clusters will be revealed. Under the framework of hypothesis-testing, we employ genes' dependence relationship as a feature to model and classify cancer and normal samples. The classification scheme is then applied to several real cancer data sets. It is noted that the method yields very promising performance. Further investigation of the eigen-value pattern of the proposed method allows us to discover that the transition in between cancer and normal patterns suggests that the eigenvalue pattern of the proposed models may have potential to predict the early stage of cancer development. A dependence network was also developed for biomarker identification and selection.

TUESDAY

Li Deng
Xiaodong He
Hermann Ney
George Tzanetakis
Dick Lyon
Les Atlas
Dong Yu

From Recognition to Understanding --- Expanding the Traditional Scope of Signal Processing

Li Deng,
Microsoft Research
Redmond

The traditional scope of signal processing as defined in the SPS constitution includes the "signal" classes of audio, video, speech, image, communication, musical, and "others", and including the "processing" classes of filtering, coding, transmitting, estimating, detecting, analyzing, recognizing, synthesizing, recording, and reproducing. In our modern information society, we immerse ourselves with the signal processing techniques and applications that go far beyond the above scope. In this talk, I construct a "matrix" which succinctly represents the traditionally defined scope of signal processing and use this matrix representation to argue for natural expansion of the signal processing scope. In particular, I argue for extension of the "signal" coverage from typical numerical type to symbolic type such as text and documents, and for extension of the "processing" class from recognition to understanding (http://research.microsoft.com/pubs/79366/Deng-SPM-editorial08.pdf).

I will present a case study which demonstrates principled ways in which the commonly used speech recognition techniques are naturally extended to handle the more challenging problem of speech understanding (with new, problem-specific processing steps added in an integrative manner).

Speech Recognition and Machine Translation: A Comparative Overview

By Xiaodong He,
Microsoft Research, Redmond

Over the last decade, substantial progress has been made in both research and real-world applications of speech recognition and machine translation. Despite conspicuous differences, many problems in speech recognition and in machine translation share a wide range of similarities, and it is of great interests to see techniques in these two fields can be successfully cross-fertilized. In this talk, I will discuss the similarities and difference between speech recognition and machine translation. As case studies, three specific technologies that have been successfully applied to both fields will be discussed in details: hidden Markov model, template based modeling and decoding, and system combination. Through these examples, we will compare the properties of speech and language, and show how generic sequential pattern recognition technologies could be extended and applied to address the particular needs of speech recognition and machine translation.
Tutorial :

Statistical Methods for image, speech, and language processing: Achievements and open problems

By Prof. Hermann Ney,
RWTH Aachen University

The last two decades have seen a dramatic progress in statistical methods for recognizing image and speech signals and for translating spoken and written language.
This talk gives an overview of the underlying statistical methods.
In particular, the talk will focus on the remarkable fact that, for all three tasks, the statistical approach makes use of the same four principles:

* Bayes decision rule for minimum error rate,
* probabilistic alignment models, e.g. Hidden Markov models,for handling strings of observations (like acoustic vectors for speech recognition and written words for language translation),
* training criteria and algorithms for estimating the free model parameters from large amounts of data,
* the generation or search process that generates the recognition or translation result.

Most of these methods had originally been designed for speech recognition.
However, it has turned out that, with suitable modifications, the same concepts carry over to both language translation and image recognition, which in both cases results in systems with state-of-the-art performance.
This lecture will summarize the achievements and the open problems in this area of statistical modelling.

Computational Ethnomusicology - Expanding the reach of Music Information Retrieval to the musics of the world

George Tzanetakis,
University of Victoria

Music Information Retrieval (MIR) is a relatively new research area in multimedia. MIR deals with the analysis and retrieval of music in digital form. It reflects the tremendous recent growth of music-related data digitally available and the consequent need to search within it to retrieve music and musical information efficiently and effectively. Most of existing work in MIR has focused on western classical and popular music as these types of music have the largest commercial interest.

In this talk I will describe two case studies in Computational Ethnomusicology that explores how MIR techniques can be applied to the study of non-Western music for which there is no standardized written reference (which is a large percentage of the music of world if not of album sales). The first case study is an automatic analysis of micro-timing in complex Afro-Cuban percussion music using rotation-aware dynamic programming. The second case study is a content and context aware web visualization interface for the study of religious chant. In addition to the technical challenges these projects presented I will also discuss the social challenges of Interdisciplinary collaborations between engineering and humanities.

Machine Hearing - A Research Agenda and an Approach

Richard F. Lyon;
Google Inc. Mountain View, CA.

The field of machine hearing is still in its infancy, in comparison with the thriving field of machine vision. This unfortunate situation, combined with the availability of good front-end auditory models, provides us the opportunity to make quick progress by leveraging techniques from machine vision to help make progress in research and applications in machine hearing. Our project at Google aims to help machine hearing become a first-class academic and commercial field. We develop applications that will do something useful with all that uninterpretable audio media out there, such as sound tracks of

amateur movies. There are three main tactics that will help us: (1) Leveraging techniques already developed in the machine-vision and machine-learning fields; (2) Productive interaction with the wider field of hearing research, to keep models honest and motivate better experiments; (3) Focus on applications for which the challenge has to do with what things sound like, as opposed to specialized domain knowledge ("non-speech non-music audio"). We have good results already, showing how very-high-dimensionality feature spaces can effectively connect auditory representations to simple but powerful machine learning techniques for a range of applications such as sound ranking from text queries.

Acoustic Scenes, Complex Modulations, and a New Form of Filtering

Les Atlas
University of Washington

Be it in a restaurant or other reverberant and noisy environment, normal hearing listeners segregate multiple sources, usually strongly overlapping in frequency, well beyond capabilities expected by current computational approaches. What is it that we can learn from this common observation? As is now commonly accepted, the differing dynamical modulation patterns of the sources are key to these powers of separation. But until recently, the theoretical underpinnings for the notion of dynamical modulation patterns have been lacking. We have taken a decades-old and loosely defined concept, called "modulation frequency analysis," and developed a theory which allows for distortion-free separation (filtering) of multiple sound sources with differing dynamics. A key result is that previous assumptions of non-negative and real modulation are not sufficient and, instead, coherent and sparse separation approaches are needed to separate different modulation patterns. These results may have an impact in separation and representation of multiple simultaneous sound streams for speech, audio, hearing loss treatment, and underwater acoustic applications. This research also suggests exciting new and potentially important open theoretical questions for general signal representations, extending beyond acoustic applications and potentially impacting other areas of engineering and physics.

Deep-structured learning in speech processing

Dong Yu,
Microsoft Research, Redmond

In this talk we report our investigations on ways to learn complexy sequential decision boundaries in speech processing by composing multiple-layers of simple learners. We show that this hierarchical structure allows us to learn and use long-range dependencies hidden in the signals and use features that cannot be easily incorporated in the hidden Markov model

WEDNESDAY

Sabine Susstrunk
Amir Said
Jeffrey Bloom

Material classification using visible and near-infrared images
Sabine Süsstrunk
EPFL

Recently, we have shown the advantages of simultaneously capturing visible and near infrared (NIR) radiation in digital photography applications, such as white balancing, shadow detection, dehazing, and face rendering. In this talk, I will present our on-going research using NIR images in conjunction with visible images for material classification. As many colorants are transparent to NIR, we are able to reproduce the intrinsic lightness and texture characteristics of a material. We are thus currently analyzing visible and NIR images according to their lightness and texture. The results are the input of a classifier in form of feature vectors, and the probability of that data to belong to a material category is then calculated. We achieve good classification results on a limited set of material classes. I appreciate the audiences feed-back and ideas on how to generalize the method to account for more materials.

Visualizing and Understanding Challenges in the Design of Light Field Displays
Amir Said
Hewlett Packard

While attempts to recreate three-dimensional views are nearly as old as photography, no solution has been able to generate consistent interest and wide acceptance of their quality. We present new analysis techniques that can easily and more naturally show why the problem is not impossible, but can be indeed very challenging. We also show sequences of display simulations, which can provide a much more intuitive appreciation of the difficulties, and facilitate understanding how design limitations impact visual quality.

Understudied Constraints Imposed by Watermarking Applications
Jeffrey Bloom

Dialogic Research Inc.

As video watermarking becomes more mature and more widely known and accepted, a number of security and non-security applications are emerging. When watermarked content is traveling through a network or series of networks, there is a need to embed and/or detect watermarks at various points in the distribution chain. Traditional watermarking research concentrated only on the input and output of the network. This leaves a number of scenarios understudied. We will discuss two such scenarios: embedding in an entropy-encoded bitstream and detection in a compressed domain that differs from the embedding domain due to transcoding.

THURSDAY

Edward Delp
Mehrdad Fatourechi
Jane Wang
Vicky Zhao
Darko Kirovski
Deepa Kundur
Ton Kalker
Yucel Altubasak
Magdy Bayoumi

Multimedia Security: A Viewpoint from a Walking Wounded

Edward J. Delp
Purdue University

This talk will describe current research issues in multimedia security. Issues involving data hiding, device forensics, biometrics, DRM, and authentication will be described. I will try to "predict" the future as to where this is all going. I hope we will have fun.
This talk will also present a brief overview of the research done in the Video and Image Processing Laboratory at Purdue University. Projects that will be described include video compression, media indexing, multimedia security, language translation, mobile applications, and medical imaging.

Information Management and Security in Media-Sharing Social Networks

Mehrdad Fatourechi
University of British Colombia

Z. Jane Wang
University of British Columbia

Hong Zhao
University of Alberta

Digital media has profoundly changed our daily life during the last decade. For example, the wide adoption of broadband residential access and recent advances in video compression technologies has fueled increasing popularity in delivery of TV services via Internet. We have also witnessed the emergence of large-scale multimedia social network communities such as Facebook and YouTube. This proliferation of digital multimedia data creates a technological revolution to the entertainment and media industries and introduces the new concept of web-based social networking communities. However, the massive production and use of digital media also pose new challenges to the scalable and reliable sharing of multimedia over large and heterogeneous networks, demand effective management of enormous amount of unstructured media objects that users create, share, distribute, link and reuse, and raise critical issues of protecting intellectual property of digital media data.

The proper management and protection of digital multimedia at such an unprecedented scale are beyond the capability of current technologies and demand new solutions. This collaborative research effort between University of British Columbia and University of Alberta tackles the emergent technical challenges (e.g. information management and content protection) in large-scale media social networks. Our aim is to establish a multimedia management and security framework to provide effective management, secure and reliable sharing of digital media in large-scale social networks. In particular, this talk addresses a summary of our recent research efforts in the following areas:

·content-based fingerprinting for media indexing and content recognition;
· understanding and analyzing the impact of human factors on multimedia systems; and
· building an automated network-service monitoring paradigm.

Realizing the Uniqueness of Optical Media

Darko Kirovski
Microsoft Research, Redmond

When a DVD is stamped it is physically unique. We propose a scheme to detect and deploy this uniqueness for the benefit of Digital Rights Management.

Connectivity and Security in Directional Multimedia Sensor Networks
Deepa Kundur

Texas A&M

Recently, there has been increased interest in the development of untethered sensor nodes that communicate directionally via directional radio frequency (RF) or free space optical (FSO) communications. Directional wireless sensor networks, such as the original Smart Dust proposal that employs broad-beamed FSO communications have the potential to provide gigabits per second speeds for relatively low power consumption suitable for multimedia sensing systems. Two significant challenges shared by the class of directional networks are connectivity and routing security, especially for random deployments. In this talk we study the feasibility of employing directional communications paradigms in large-scale security-aware broadband randomly and rapidly deployed static multimedia sensor networks. We investigate the implications of link directionality to network connectivity and secure ad hoc multihop routing and highlight approaches in network design to mitigate compromising between the two.

Bounds on Biometric Security

Ton Kalker
Hewlett Parkard

Abstract: We give an overview of some recent work on trade-offs between the capacity and security of biometric systems. We show that it is possible to formulate bounds for a number of cases, and that some of the classical schemes (for example fuzzy commitment) are sub-optimal. We also sketch some of the many open questions.

The H- INDEX : DISCUSSION

Yucel Altunbasak
Georgia Tech

A while ago, with the help of some Ph.D. students, I compiled H-indices for image processing researchers. I received several feedback about publishing this list. The most common feedback were:
1) How do you define an image processor?
2) How do you resolve the different H-index characteristics of image processing, computer Vision, and computer-science oriented people?
3) ISI-based and Google-based H-index numbers are vastly different. Which one do you use?, and most important of all,
4) what is the use of publishing such a list? Would it benefit the society?

I still need to collect further feedback from the society regarding 1) if publishing such a list (online) would benefit the society or be harmful because of possible misinterpretation (esp. by students), and 2) how best to do this project.

This is not a panel discussion per se .It is an interactive discussion. I will just describe the project and ask several questions to the attendees in an effort to get as much feedback as possible

We describe a technique for using dual-tree complex wavelets to obtain rich feature descriptors of keypoints in images. The main aim has been to develop a method for retaining the full phase and amplitude information from the complex wavelet coefficients at each scale, while presenting the feature descriptors in a Fourier-domain form that allows for efficient correlation at arbitrary rotations between the candidate and reference image patches. The feature descriptors are known as Polar-Matching matrices. Recently, we have modified our previously proposed approach so that it can be more resilient to errors in keypoint location and scale. These multi-scale feature descriptors are
potentially useful for object detection, recognition, classification and tracking in images and video.

Cognitive Sensors Networks: The New Frontier for DSP
Magdy Bayoumi
University of Louisiana at Lafayette

Computers, communication, and sensing technologies are converging to change the way we live, interact, and conduct business. Wireless sensor networks reflect such convergence. These networks are based on collaborative efforts of a large number of sensor nodes. They should be low-cost, low-power, and multifunction. These nodes have the capabilities of sensing, data processing, and communicating. Sensor networks have a wide range of applications, from monitoring industrial facilities to control and management of energy applications, to military and security fields. Because of the special features of these networks, new network technologies are needed for cost effective,low power, and reliable communication. These network protocols and architectures should take into consideration the special features of sensor networks such as: the large number of nodes, their failure rate, limited power, high density, etc.
Moreover, applications and impact of Sensors Networks are going to a higher and wider levels through the development of cognitive capabilities of these networks. Cognitive Sensors Networks, CSN, represent a transformational impact on technologies, applications, and expectations. In this talk the impact of wireless sensor networks will be addressed, several of the design and communication issues will be discussed, and a case study of a current project of using such networks in drilling and management off-shore oil and natural gas in the gulf region will be given. The main criteria, expectations, and objectives of

CSN will be also highlighted