Multimedia and Mathematics

Rabab Ward (Institute for Computing, Information and Cognitive Systems, UBC),
Robert Gray (Stanford University)

July 23, 2005–July 28, 2005

1 Introduction

The diverse applications of multimedia technology affect the way we communicate, work and play. The Banff International Research Station (BIRS) workshop on Multimedia and Mathematics, organized by Rabab K. Ward and Robert Gray, brought university and industry personnel together from July 23-28, 2005 to share ideas about the latest advances in the different areas of multimedia and related mathematics. Forty attendees (29 men and 11 women) from Canada, UK, Australia and the USA comprised 6 graduate students and 26 faculty members from 24 universities, as well as 8 researchers from Microsoft, Apple, Hewlett-Packard, Tiz Media Foundation, and the National Science Foundation. The rich cross-fertilization brought about by this workshop provided new insights into possible solutions to the latest technical challenges.

Academics and mathematicians, as well as practitioners, engineers, and researchers working in different industries related to multimedia devices, described the approaches, advances, and constraints involved in their area of media. With a view to discovering common ground, they explored the mathematical modeling, analysis, and representation of information relevant to their respective fields. Models used in individual media as well as in multimedia systems were examined. Under this broad umbrella, the following topics were discussed: algorithms, architecture and hardware, software, joint processing and coordination of multi-model signals and data, coding, compression, storage, retrieval, statistical learning, recognition, classification, segmentation, communication, networking, multi-model devices and systems, multimedia forensics and security, human movements and mobile devices.

The main types of signals discussed in the workshop were those involved in text, audio, speech, music, images, and video, as well as sensor data such as environmental measurements from sensor networks and biological data from medical devices. The role of multimedia in hip-hop culture was also investigated as a means of promoting mathematics among under-represented minorities.

Among the many topics discussed, three important areas received special attention: (1) data protection; (2) coding; and (3) reduction in the computational load of multimedia devices and processors. Multimedia networking and security are intertwined topics because the growth of multimedia products raises concerns for content producers about how best to protect their information. At the same time, there is a need to make better use of bandwidth in a network hardware infrastructure whose standards are fixed. This need for greater bandwidth efficiency is reflected in the number of presentations in the coding area. Reducing computational complexity received much attention, since future multimedia communication will be based on wireless devices with person-to-person connections.

Peer-to-peer video streaming and wireless multimedia represent a paradigm shift. Traditionally, most video content has originated from only a few places (mainly broadcasters) for mass distribution to consumers. Now, however, consumers equipped with digital cameras, camcorders, and camera phones, have multiple ways of generating, acquiring, and managing their own video content. Video now comes from a multitude of sources, and not a lot of computing power can be crammed into these mobile imaging devices without draining their batteries and using up their limited data storage. Along with the limited bandwidth of wireless devices, this limitation requires that the video signals be compressed. However, functions such as motion estimation and compensation, which are integral to video compression, are very computer-intensive. Today, compression and other video encoding are done by broadcasters. For mobiles, however, we need to shift the computationally demanding components such as motion estimation and compensation to the desktop machine or the mobile. Mobiles, therefore, need to have simpler, less power-hungry encoders, and we need to reduce encoder complexity in ways that wont affect compression efficiency. We arent really there yet.

In the following two sections, we summarize the topics explored at the workshop. For convenience of presentation, we classify the topics discussed into categories, ”Theory and Modeling” and ”Progress in Specific Application Areas”, even though almost every presentation discussed theory as well as applications.

2 Presentations and Discussion

2.1 Theory and Modeling

Of the 25 presentations, 13 can be roughly categorized under Theory and Modeling. In most of these talks, different applications to multimedia applications were also discussed and demonstrated. The following four talks could fall under the general topics of modeling images, image rendering, humancomputer interaction and a unified algebraic approach to time and signal models: Photographic Image Representation with Multiscale Gradients and Applications, e.g., to Denoising Taking Multi-View Imaging to a New Dimension: From Harry Nyquist to Image-Based Rendering A New Framework for Modeling and Recognizing Human Movement and Actions Deterministic and Stochastic, Time and Space Signal Models: An Algebraic Approach The area of image and video coding received much attention, as mentioned earlier. The following four talks were given in this area: A Signal Processor’s Approach to Modeling the Human Visual System, and Applications, e.g., to Coding Vector Quantizers for Reduced Bit-Rate Coding of Correlated Sources Analytical Modeling of Matching Pursuit Time Domain Lapped Transform and Its Applications to Coding and Error Resilience Transmission Additional presentations in this area are discussed in the following section on Progress in Specific Application Areas. There was one presentation on information representation of networks, entitled Information Representation for Network Systems. Image segmentation remains an active area of research, with many applications ranging from video retrieval to biomedical imaging. The following two presentations addressed image segmentation: Mathematical and Perceptual Models for Image Segmentation Deformable Models for Image Analysis: From ’Snakes’ to ’Organisms’ Two presentations that addressed reduction in computational load were Dimension Reduction for Classification and Anomaly Detection Multi-scale Displacement Estimation and Registration for 2-D and 3-D Datasets. We will now briefly describe the digests of the above talks, highlighting recent developments, scientific progress and some of the open problems.

Eero Simonelli talked about photographic image representation with multiscale gradients. He described recent empirical investigation and modeling of the joint statistical properties of a multiscale representation based on derivative operators. In particular, he discussed the use of Gaussian Scale Mixtures (product of a scalar random variable and a Gaussian vector) to model the statistics of clusters of wavelet coefficients at adjacent positions, scales and orientations. When applied to the problem of denoising, these models provide a natural generalization of both standard linear (Wiener) and thresholding estimators, and lead to substantial increases in performance. He also described how to extend this model to include local geometry in the form of phase and orientation information.

Tsuhan Chen talked about the recent convergence of image processing, computer vision, and computer graphics resulting in multi-view image processing. A picture may be worth a thousand words, but a single picture is not able to render the whole scene; it merely renders the scene as seen from a particular viewpoint. In 1991, Adelson and Bergen proposed the concept of the plenoptic function, a seven-dimensional function that represents all the light rays in a dynamic scene. Since then, research on sampling, storing, interpolating, and reconstructing the plenoptic function has been emerging at both academic and industrial research institutions. This area of research is commonly referred to as image-based rendering, or, more familiar to the signal-processing community, multi-view image processing. Recent convergence of image processing, computer vision and computer graphics has resulted in significant progress in multi-view image processing. Now widely used in applications ranging from special effects (remember the movie ”The Matrix”?) to virtual teleconferencing, multi-view image processing has become a critical tool for creating visually exciting content. With multi-view image processing, real-world scenes can be captured and rendered directly from images captured by cameras, eliminating the need for computationally expensive modeling of 3D geometry or surface reflectance, as is often done in traditional computer graphics. Dr. Chen also discussed recent developments in image-based rendering. While studying the mechanism for sampling multi-view data, he revealed the connections between image-based rendering, multidimensional multirate signal processing, and the Sampling Theorem discovered by Harry Nyquist 80 years ago!

Ling Guan described a new framework for modeling and recognizing human movement and actions. Humancomputer interaction (HCI) study is a key research area in many scientific disciplines. Dr. Guan started the talk with an overview of concepts, history and recent developments in HCI: face, speech, gesture, human emotion and human actions, with emphasis on emotion and action recognition. He then focussed on a fundamental, but under-investigated research area in HCI: modeling and recognizing human movement and actions. Inspired by the movement notation systems used in dance and the paradigm of the phonemes used in continuous speech recognition, he described a Continuous Human Movement Recognition (CHMR) framework. The framework is based on a novel paradigm, the alphabet of dynemes, the smallest contrastive dynamic units of human movement. A Differential Evolution-Monte Carlo particle filter is introduced, which has demonstrated highly effective and robust characteristics in tracking basic human movement skills. Using multiple hidden Markov models, the recognition process attempts to infer the human movement skill that could have produced the observed sequence of dynemes. Recent anthropometric data shows that the famous ”average sized human” model in Leonardo da Vinci’s drawing of the human figure is a fallacy, and that there is no one who is average in 10 dimensions. Incorporating the highly accurate biometric features into the CHMR framework, Dr. Guan was able to demonstrate the effectiveness of the framework in biometrics, biomedical analysis, and recognition of human skills. He proposed and forecasted that this framework will form the enabling technology for biometric authentication systems for a broad range of applications such as security/surveillance, biomedicine/physiotherapy, special effects in motion picture production, digital asset management, battlefield surveillance, coaching/training/judging in sports and performing arts, to name a few.

Jose Moura presented a new algebraic approach for deterministic and stochastic, time and space models. We are all familiar with (infinite) ”time” signal processing: time shifts, filters and convolution, signals, Fourier and z-transforms, spectrum, fast algorithms. Images, of course, are not ”time” but ”space” objects. Also, they are ”finite” objects, i.e., defined over a finite indexing set. What is the natural concept of space shift, of space filter and convolution, spectral analysis, or ”z”-transform, as well as many other related concepts? To address these questions, Dr. Moura went beyond linear algebra to present an algebraic approach where time (signal) and space (image) processing are instantiations of the same mathematical structure. The basic building block is the signal model - a triplet (A, M, f) of an algebra A of filters, a module M of signals, and a generalization of the z-transform as a bijective linear mapping f from a vector space into the module of signals. The shift is naturally interpreted as a generator of the algebra of filters, boundary conditions connect finite with infinite indexing sets, the trigonometric transforms (e.g., DCTs) are appropriate Fourier transforms, and the C-transform is the z-transform. More than a mathematical curiosity, the algebraic approach provides the appropriate structure to extend signal and image processing beyond uniform to other grids (e.g., hexagonal or quincunx), or develop fast algorithms from a few basic principles, from which we can also derive new fast algorithms for existing and new transforms. Connections with other image models, in particular, with Gauss Markov fields and pinned Markov diffusions, were discussed. This talk overviewed Moura’s recent work with Markus Pueschel on the algebraic theory of signal and image processing.

Sheila Hemami presented a signal processing approach to model the human visual system. Current image and video compression algorithms (e.g., JPEG-2000, H.264) provide very high efficiency compression and excellent quality at relatively high bit rates. These algorithms operate by treating images and video as traditional ”signals,” employing efficient transformations, correlation-based models, and entropy coding. Human visual system characteristics have been successfully applied to high-rate signal-based compression, where stimuli such as compression-induced distortions are below the visibility threshold; i.e., humans cannot see them. Operation of such signal-based compression algorithms at low rates, in which compression-induced distortions are clearly visible, has to date operated based on visual system rules-of-thumb and has produced moderate success for images, while little has been done for video. Dr. Hemami presented recent results on characterizing the human visual system in a manner that allows for immediate incorporation into imaging and video applications, such as compression and quality measurement, at not only high rates/low distortions but also at low rates/high distortions. Results were presented in two distinct areas: vision-based results that explain how humans perceive stimuli, and engineering-motivated results that allow us to incorporate our characterizations into practical algorithms.

Russ Mersereau discussed coding of correlated sources. It is well known that vectors derived from consecutive segments of most real-world signals are strongly correlated. This inter-vector correlation is not exploited in a standard VQ system. Many techniques proposed to exploit this correlation render the VQ sub-optimal or require buffering, and thus introduce encoding delay. Dr. Mersereau presented two alternative methods. The first approach, cache VQ, uses a cache memory to reduce the bit rate and the encoding time, at the cost of a slight, but controllable, increase in the coding error. The second approach, recently developed by Krishnan, Barnwell, and Anderson at Georgia Tech, overcomes cache VQ's limitations. Their approach, called dynamic codebook reordering, dramatically reduces the entropy in the representation of the VQ symbols, which can then be exploited for lossless compression. Dynamic codebook reordering can significantly reduce the bit rate for strongly correlated sources without introducing any additional distortion, coding delay, or sub-optimality when compared to a standard VQ.

Shahram Shirani presented an analytical approach that models the operation of the matching pursuit algorithm on uniformly distributed signals. Matching pursuit is a greedy algorithm that decomposes a signal into a redundant dictionary of basis functions. It has recently found applications in many areas, including image and video processing. The proposed model expresses the relationship between the bit rate and matching pursuit coder parameters such as dictionary size, quantization step size, distortion and dimension of the signal. This relationship can be used to optimize the dictionary size and quantization step size for minimum bit rate. The model is verified through experimental results, and the accuracy of the model is validated for different system parameters.

Jie Liang reviewed the theory and applications of time domain lapped transform, including the design of fast transform, its application in wavelet-based image and video coding, and error resilient design for multiple description coding.

Thrasos Pappas discussed problems arising in the segmentation of images of natural scenes. One of the challenges of this problem is that the statistical characteristics of perceptually uniform regions are spatially varying due to effects of lighting, perspective, scale changes, etc. A second challenge is the extraction of perceptually relevant information. Dr. Pappas first considered the problem of segmenting images of objects with smooth surfaces. The images are modeled as smooth spatially varying functions with sharp discontinuities at the segment boundaries, plus white Gaussian noise. Dr. Pappas discussed an adaptive clustering algorithm for segmentation, which is a generalization of the K-means clustering algorithm to include spatial constraints and to account for local intensity variations in the image. The spatial constraints are modeled through the use of Gibbs/Markov random fields, while the local intensity variations are accounted for in an iterative procedure involving averaging over a sliding window whose size decreases as the algorithm progresses. Dr. Pappas also considered a hierarchical implementation that results in better performance and computational efficiency, then discussed an adaptive perceptual colortexture segmentation algorithm that is based on low-level features for color and texture. It combines knowledge of human perception with an understanding of signal characteristics in order to segment natural scenes into perceptually/semantically uniform regions, and is based on two types of spatially adaptive low-level features. The first describes the local color composition in terms of spatially adaptive dominant colors, and the second describes the spatial characteristics of the gray-scale component of the texture. Key segmentation parameters are determined on the basis of subjective tests. The resulting segmentations convey semantic information that can be used for content-based retrieval.

Another presentation on image segmentation was given by Ghassan Hamarneh. Dr. Hamarneh started by giving a short overview on image segmentation and registration. He then focussed on deformable models ('snakes' and others) for image segmentation and mentioned issues related to incorporating prior knowledge. He then presented his work on 'deformable organisms', an artificial-life framework for image analysis incorporating high-level, intelligent, intuitive control of shape deformations. Various application examples were presented throughout the talk.

Dimension reduction for classification was discussed by Alfred Hero. There has been intense interest in analysis of massively complex data sets with thousands of dimensions. Dimension reduction methods are critical components of any analysis method due to the requirements of computation and noise reduction. Dr.

Hero presented new variational methods of dimension reduction that explicitly target classification, anomaly detection, or other tasks.

Nick Kinsbury discussed the problems in motion estimation and registration of images and 3-dimensional objects. His talk considered the problems of displacement (or motion) estimation between pairs of 2-D images or 3-D datasets, especially for the case of non-rigid deformation as encountered in many medical imaging applications. He showed how the use of multi-scale directionally selective octave-band filters with analytic (complex) impulse responses can greatly reduce the computational load associated with displacement estimation by employing phase-based methods. In particular, he extended the techniques of Hemmendorf for use with dual-tree complex wavelets (DT CWT) and in an iterative scenario, such that the usual approximations associated with phase-based approaches are minimized. These methods rely on the shift-invariant and directional properties of the DT CWT, and are inherently resilient to shifts in the mean level and contrast of the two datasets and to noise, because of the band-limited nature of the signals and the use of phase shifts to estimate displacements. They are computationally efficient because a coarse-to-fine, multi-scale approach is used, and they are well-suited to displacement fields that can be represented by locally-affine models with smoothly varying parameters. The algorithm can also be designed largely to ignore data in areas where the two datasets do not match (e.g., where a tumour is present in one dataset but not in the other). Dr. Kinsbury believes that the computational advantages of this method will be particularly helpful for 3-D registration tasks.

2.2 Progress in Specific Application Areas

The areas of forensics and security, video coding, automated speech recognition, automated music retrieval, video for mobile devices and network coding for the Internet and wireless networks were discussed. A presentation of a different kind but which received much discussion was that of using multimedia and hip-hop culture to promote math among under-represented minorities. There were three presentations on forensics and security, entitled Multimedia Forensics for Traitors Tracing Secure Signal Processing Emerging Paradigms in Sensor Network Security Dr. Adrian Dumitras of Apple Inc. and Dr. Amir Said of Hewlett Packard talked about the state of the art in video coding. The titles of their presentations were Optimization Methods for State-of-the-Art Video Encoders The Need for Better Models for Coding Sparse Multimedia Representations Workshop attendees also discussed recent developments and open problems in the area of speech and music. The following three talks addressed this field: Computer Speech Recognition: Building Mathematical Models Mimicking the Human System Managing Spoken Documents A Personal History of Music Information Retrieval Panos Nasiopoulos and Kostas Plataniotis gave a joint presentation regarding consumer-grade mobile devices. The titles of these presentations were Digital Video for Mobile Devices A Unified Framework for the Consumer-Grade Image Pipeline Philip Chow of Microsoft gave the following talk on the newly emerging theory and applications of network coding: Network Coding for the Internet and Wireless Networks

Ray Liu presented first on the art of multimedia security. The recent growth of networked multimedia systems has increased the need for techniques that protect the digital rights of multimedia. Traditional protection alone (such as encryption, authentication and time stamping) is not sufficient for protecting data after it is delivered to an authorized user or after it has traveled outside a closed system. To address the post-delivery protection and introduce user accountability, a class of technologies known as digital fingerprinting is emerging. Due to the global nature of the Internet, ensuring the appropriate use of media content is no longer a traditional security issue with a single threat or adversary. Rather, new threats are posed by coalitions of users who can combine their contents to undermine the fingerprints. These attacks, known as collusion attacks, provide a cost-effective method for removing an identifying fingerprint, and thus pose a strong threat to protecting the digital rights of multimedia. To mitigate the serious threat posed by collusion, theories and algorithms are being investigated and developed for constructing forensic fingerprints that can resist collusion, identify colluders, and corroborate their guilt. Therefore, multimedia forensics has become an emerging field built upon the synergies between signal processing theory, cryptology, coding theory, communication theory, information theory, game theory, and the psychology of human visual/auditory perception. Dr. Liu provided the audience with a broad overview of the recent advances in multimedia forensics, with a focus on multimedia fingerprinting for traitor tracing. He then talked about tracing traitors using collusion-resistant fingerprinting for multimedia that jointly considers the encoding, embedding, and detection of fingerprints.

A general formulation of fingerprint coding and modulation with a unified framework covering orthogonal fingerprints, coded fingerprints, and group fingerprints was discussed. Finally, traitor-within-traitor dynamics and behavior was modeled and analyzed. As a result of this work, optimal strategies for traitors and for detectors can now be developed.

Ton Kalker talked about secure signal processing. He observed that (professional) multimedia signals are increasingly made available only in protected format. Typically, the security wrappers can only be removed by the targeted devices or applications (e.g., the DRM agent in a rendering device). This poses serious problems for intermediate processing applications that do no have access to the appropriate cryptographic keys (for liability reasons, security reasons or otherwise) and/or that do not have sufficient computational resources. In his talk, Dr. Kalker discussed options for processing of protected signals in their protected format, both by adopting the cryptographic methods (e.g., homomorphic encryption) or by adapting the signal processing methods (scalable coding).

Deepa Kundur talked about the emerging paradigms in Sensor Network Security. She provided an overview of the field of sensor network security and highlighted particular challenges in symmetric key distribution, secure aggregation, secure routing, and actuation security. Through examination of these problems, fundamental compromises among the degree of protection, complexity and network performance were highlighted, leading to a discussion of appropriate primitives and paradigms for securing sensor networks. The talk concluded with a discussion of the principal issues for protecting emerging optical free space sensor networks and multimedia sensor networks.

Adriana Dumitras discussed optimization of video encoders. Much work has been done on identifying the best methods to optimize video encoders. These efforts have focused on removing spatial, temporal and perceptual redundancies from a video source, with the objective of representing the data efficiently. However, so far there is no unique "best method" to optimize a video encoder. Instead, various methods exist that address (usually distinctly) different aspects of the optimization problem and different applications. This diversity is motivated and enabled by the tremendous flexibility allowed in the encoder design by video coding standards, the development of unoptimized video encoding tools as part of the non-normative verification or experimental models in the standards' developments, and the powerful competition in the video industry. Dr. Dumitras presented a taxonomy and an overview of the methods that enable video encoder optimization by tradeoffs at the algorithmic, software and hardware implementation levels.

Amir Siad talked about the need for better models for multimedia coding. A main objective in multimedia signal processing is to numerically eliminate redundancy and create sparse representations. However, for compression an effective representation needs to be effectively entropy coded. There is a need to have good combined models for both the signal and how its information is distributed, in the sense of what and where the most important components are. Simple recursive set-partitioning methods were shown to be very effective in coding sparse data, both in terms of compression and computational complexity, but their use still has not been extended to more complicated media types. Dr. Said discussed the challenges and possibilities for improving performance using more sophisticated data models.

Li Deng of Microsoft discussed computer speech recognition and how to build mathematical models that mimic the human system. The main goal of computer speech recognition/understanding is to automatically convert naturally uttered human speech into its corresponding text (and then into its meaning). While amazing success, both technologically and commercially, has been achieved in the past by straightforward mathematical methods (e.g., hidden Markov modeling, maximum likelihood and discriminative learning, dynamic programming, etc.), solving the remaining problems leading to its ultimate success appears to require a deep understanding of human speech recognition mechanisms. Dr. Deng analyzed various human sub-systems, including linguistic-concept generator, motor-control, articulation, vocal tract acoustic propagation, ears, auditory pathways, and auditory cortex, working in synergy to accomplish the remarkable task of highly robust, low-error speech recognition/perception and understanding. How can the essence of such human information processing power be abstracted in building a computer system with similar (or better) performance? How can we build mathematical models to enable the development of advanced machine-learning algorithms and techniques that will run efficiently in a computer? Is it possible to explore and exploit some special power of the computing machines inherently lacking in the human system so as to achieve super-human speech recognition? These are some of the issues Dr. Deng addressed in the talk.

Mari Ostendorf talked about the management of spoken documents. As storage costs drop and bandwidth increases, there has been a rapid growth of information available via the Web or in online archives, raising problems of finding and interpreting collections of documents. Significant recent progress has been made in text retrieval, analysis, summarization and translation, but much of this work has focused on written language. Increasingly, speech and video signals are also availableincluding TV and radio broadcasts, congressional records, oral histories, voice mail, call center recordings, etc.which can be thought of as spoken documents'. Because it takes longer to listen to audio than to read text, spoken documents are clearly a prime candidate for automatic indexing, information extraction, and other such technologies. In her talk, Dr. Ostendorf provided an overview of the speech processing technology underlying spoken document management, including mathematical frameworks for both word and metadata recognition, and for integrating video and language cues. In addition, she discussed issues that arise in text processing when moving from written to spoken language and implications for statistical models of language.

George Tzanetakis gave a very lively presentation about music retrieval, complete with beautiful and varied pieces of music. Music Information Retrieval (MIR) is an emerging research area that explores how large digital collections of music can be effectively analyzed for searching and browsing. It combines ideas from many different fields, including Signal Processing, Machine Learning, Music Cognition, and Human-Computer Interaction. Dr. Tzanetakis gave a historic overview of MIR, with specific emphasis on topics he had more personal experience with, such as audio-feature extraction, automatic musical genre classification, rhythm analysis, query-by-humming, and sensor-enhanced musical instruments. He concluded the talk by making predictions about the future of MIR and how it will radically transform the way music is produced, distributed and consumed.

Panos Nasiopoulos talked about digital video and mobile devices. Mobile wireless technologies and digital video broadcast technologies are gradually converging within efforts from 3GPP and DVB 2.0 to complete this merging in the upcoming generations of mobile technologies. In order to support this convergence, existing video technologies need to be upgraded to ensure the reliability and quality of the delivered content. This calls for highly efficient video codecs in addition to reliable error resilience techniques that overcome the bandwidth constraints and highly error-prone conditions of wireless networks.

Kostas Plataniotis talked about a unified framework for consumer grade image pipeline. A new modeling and processing approach suitable for consumer-grade image processing was presented. Using vector modeling principles, nonlinear image operators and adaptive filtering concepts, single-sensor camera image processing problems are treated from a global viewpoint yielding new classes of processing solutions. The following varied applications of the framework were covered: spectral interpolation (demosaicking), spatial interpolation of the acquired (mosaic-like) single-sensor gray-scale images as well as demosaicked full-color images, demosaicked image post-processing and color image enhancement, camera image denoising and sharpening, camera image compression, spatio-temporal video demosaicking, and camera image indexing and rights management. Results obtained using the framework were provided. The list of the topics covered, while certainly not exhaustive, provided a good indication of the usefulness and often necessity of the proposed framework in consumer grade image processing. Open research problems and other potential applications of the framework were also discussed.

Melanie Louisa West gave a multimedia presentation about using multimedia and hop-hop culture to teach math to under-represented minorities. There is widespread agreement among educators that a strong need exists for programs to increase math and science competency among under-represented minority students. Lack of interest and motivation are known contributing factors for this lack of representation. Ms. West proposed combining multimedia with elements of hip-hop culture to promote interest in math among under-represented minorities. In today's society, hip-hop music has captured the minds of urban youth. Music sales, fashion trends, and advertisement strategies reflect this. Consequently, Ms. West believes that incorporating hip-hop into math instruction for under-represented minorities holds great promise for success. The elements of rhyme, rhythm, and repetition make raphip-hop's linguistic componentan excellent creative vehicle for presenting concepts that require memorization. Math, in particular, lends itself to rap because the creative use of natural language provides a platform for transferring the conceptualization of math into real life experiences through story-telling. By combining the learning experience with an activity that is already an integral part of a person's life, Ms. West believes that this will not only increase interest in learning, but will also maximize information retention. This coupled with the incorporation of multimedia elements that will be widely accessible (on public display for peers and or publish for a general audience) will motivate the individual (or group) to do their very best. An innovative aspect of this proposed approach is that it combines teaching students at the elementary school level with multimedia content created by students at the high school level. This accomplishes two goals; it makes it easier to motivate the younger students, and at the same time, provides a great vehicle for exposing the older students to multimedia.

3 Summary and Highlights

The workshop topic provided a timely cross-disciplinary bridge between the relatively new area of multimedia and the well-established discipline of mathematics. For many researchers in a specific area of multimedia, the workshop provided an excellent opportunity to broaden their perspective. The workshops high-quality presentations made clear the surprisingly similar mathematical approaches applied to speech, audio, image, and video-processing research.

The presentations and informal discussions enabled participants to examine the variety of approaches in different media areasan invaluable opportunity made possible by the mixed formalinformal style of the workshop. For example, the group discussion resulting from Amir Said 's presentation confirmed that coding can only be optimized if we have good models. Another example is that of Professor Pappas' presentation on image segmentation, which generated heated debate by questioning the need for an intermediate step, given that the final task is semantic image understanding/classification. The researchers with speech recognition/understanding expertise have found that integrated pattern-recognition approaches that avoid the step of speech segmentation always provide better results than modular processing approaches that involve explicit segmentation. The discussions on such disparities provided much needed information that will hopefully generate new interest in cross-media research and exploration. Li Deng, the General Chair of the 2006 IEEE Workshop on Multimedia Signal Processing, was one of the attendees. He decided, together with the Technical Committee, to continue such discussions and explorations with a special panel at the upcoming workshop on ”Differences and Similarities of Image/Video and Speech/Audio Processing Techniques.” Professor Pappas has accepted their request to organize the panel. We believe that this will have a significant impact on the future of multimedia research, an initiative inspired by this BIRS workshop.

We hope BIRS will continue sponsoring cross-disciplinary workshops such as the one we organized. Cross-disciplinary research sharing similar mathematical approaches stands to benefit the most from such workshops. The different branches of media processing research make it impossible to gain expertise in every sub-area, and this BIRS workshop helped immeasurably to foster an awareness of new trends in the various sub-disciplines. This is particularly important to some industrial researchers whose work has a relatively short-term scope. Most researchers in multimedia cannot afford the time-consuming process of mastering the subtleties of all the multimedia processing techniques. The BIRS workshop provided an ideal opportunity to make close connections among them and to deepen our understanding of problem areas.

The workshop succeeded in its aim to bring mathematicians, engineers, and scientists to interact and get exposed to each others' ideas and advances in these disciplines. As different multimedia technologies have evolved and continue to evolve at a very rapid rate, the exact definition of multimedia remains illusive, even though multimedia technologies are now being widely deployed in industries in a multitude of applications. All of these applications affect the way we live, communicate, interact with each other, work, and play.

The cross-fertilization among the different disciplines, academics and practitioners, engineers and mathematicians encouraged by the workshop was very useful in exposing the different communities to a new range of challenging and timely technical advances, the underlying mathematical problems and applications, and implementation challenges.