ACM SIG Multimedia Strategic Retreat

Participant Position Papers

October 31 – November 1, 2003
Berkeley, CA USA

Table of Contents

The following participants did not submit a position paper

- Sid Ahuja
- Marc Davis
- Ramesh Jain
- Wolfgang Klas
- Dwight Makaroff

Background

The idea for this strategic retreat first occurred at ACM Multimedia 2001 (MM'01) in Ottawa, Canada. At the time discussions by conference participants questioned the relevance and direction of multimedia research being pursued by the community. The annual conference is the premier multimedia conference as evidenced by the quality of papers published and the quantity of papers submitted for consideration each year. While the conference has been successful, it has not generated the attendance numbers or impact some people think it should be generating. We need to discuss the direction of the field, specifically, what are the grand challenges that we should be attacking and how will multimedia research impact real users?

The multimedia field is inherently interdisciplinary. In recognition of this fact, the annual conference is divided into three tracks (i.e., applications, content, and systems). Some people say the conference is essentially three separate conferences held in the same location at the same time, rather than being a coordinated activity. Moreover, few researchers identify multimedia as their primary research field. People identify their field as, among others, computer networks, signal processing, user interfaces, or databases. The interdisciplinary nature of the field may be a positive attribute but we must continue to examine, question, and understand the fundamental concepts and principles that identify multimedia as a distinct field.

Many issues relating to SIGMM operations need to be discussed too. First, the annual conference is the premier conference for multimedia research as mentioned above. Some people have suggested that we enlarge the conference to include more mini-conferences held in parallel. Other people have suggested that we develop more special topic workshops and conferences that will be held at other times of the year. Second, the Multimedia Systems Journal (MMSJ) published by Springer-Verlag, successfully rejuvenated by Klara Nahrstadt, is not treated by ACM as a first-class journal. MMSJ is not included in the ACM Digital Library. Many people have asked why it is not included? Third, SIGMM has recovered financially from the disastrous MM'96 and MM'97 conferences. We have a stable model for covering expenses, including the required ACM contribution to SIG Services, and generating a small profit each year. Our fund balance is above the required threshold, given the number of members (roughly 500). Nevertheless, we must still be careful about expenses. We need to discuss possible ways to generate more income so that we can initiate more programs for the community.

After discussing these issues in the SIG Multimedia (SIGMM) Executive Board, we decided a small workshop that invited senior members of the research community and industry should be convened to discuss the multimedia field, research directions, and SIGMM operations. This document contains short position papers from the retreat participants produced before the meeting.

Ramesh Jain, *Georgia Institute of Technology*
Lawrence A. Rowe, *U of California, Berkeley*
2003 SIGMM Retreat Organizers

Retreat Program

Multimedia Research Directions (Friday Oct 31)

8:30 – 9:30	**Welcome and Introductions**
9:30 – 10:00	**Research Viewpoints - 1** Nicolas Georganas (U of Ottawa) Sid Ahuja (Lucent) Shi-fu Chang (Columbia)
10:00 – 10:30	**Morning Break**
10:30 – 11:00	**Research Viewpoints – 2** Lynn Wilcox (FX Pal) Dick Bulterman (CWI) Tat-Seng Chua (National U of Singapore)
11:00 – 12:00	**Group Discussion**
12:00 – 1:30	**Lunch**
1:30 – 3:00	**Breakout Sessions**
3:00 – 3:30	**Afternoon Break**
3:30 – 5:00	**Group Discussion**
6:00 – 9:00	**Dinner** – Spenger's

SIGMM Future Directions (Saturday Nov 1)

8:30 – 9:00	**Further Thoughts on Multimedia Research Directions**
9:00 – 10:00	**ACM and SIGMM Report** Larry Rowe (UC Berkeley)
10:00 – 10:30	**Morning Break**
10:30 – 12:00	**Breakout Sessions**
12:00 – 1:30	**Lunch**
1:30 – 3:00	**Group Discussion**
3:00 – 3:30	**Afternoon Break**
3:30 – 5:00	**Group Discussion?**

Brian Bailey (U of Illinois Urbana Champaign)

1. Fundamental Topics and Areas

I believe that multimedia research is comprised of four main research areas:

- Content analysis and processing. This research area includes methods for coding/decoding content, encrypting content, summarizing and adapting content, and extracting semantics from content.
- Content retrieval. This research area includes methods for indexing dynamic content and for performing content-based queries of multimedia data.
- Networking and systems. This research area includes investigating quality of service issues, adaptive transmission and playback, multicast schemes, and middleware systems.
- Tools and applications. This research area includes authoring and design tools for multimedia applications and the use of multimedia in education, entertainment, storytelling, and more.

In coming years, I believe that these areas will continue to be the fundamental areas in multimedia research, but that emphasis will slowly shift to the latter research area because as more users are able to produce high-quality digital content, there will be an increasing demand for more effective authoring and design tools, which once created, should better enable users to build novel multimedia applications.

2. Successes of Multimedia and Solved Problems

The multimedia community has enjoyed many successes, but the more notable successes include:

- Enabling transparent delivery of dynamic content over the Internet. Because of research successes in compression, networking, playback, and more, end users are generally able to view dynamic content over the Internet. The success lies in the fact that most users now perceive this as "trivial." A user currently needs little knowledge of the underlying complexity of how media is coded, transmitted, or played back to use it, causing the technology to fade into the background of the interaction experience. Technology pioneers can often claim success once users no longer have to fully understand how the technology works in order to use it.
- Enabling users to author interactive multimedia applications for the World Wide Web. Through standardized markup languages such as SMIL, users have been empowered to publish interactive multimedia applications on the World Wide Web. This capability combined with the emerging ubiquity of content generation tools may revolutionize how users tell stories in the future.

3. Research Problems

The problems that the multimedia research community should focus more on include:

- Helping designers design more effective multimedia applications. Applications may include educational multimedia applications, interactive TV/DVD applications, interactive games, interactive media art, or multimedia stories. To develop a higher-quality multimedia application faster, a designer needs more effective tools in the early stages of design. These tools should enable designers to better explore and communicate interaction and media design before committing to a particular design alternative. In particular, designers need more effective tools to help design interactive animation as part of the broader interactive design. In the coming years, there will be a growing research emphasis on developing and evaluating technology that better supports human creativity.
- Helping end users design multimedia stories. With increasing use of digital camcorders, digital cameras, and MP3 technologies, more end users will have the ability to produce high-quality media content. Ostensibly, end users will want to compose meaningful stories from this content, turning every end user into a multimedia designer. The complexity of today's authoring tools, however, makes them inaccessible to most end users. I believe that recent work on interactive sketching tools may provide a better mechanism that end users can utilize to design and author multimedia stories.
- Helping users at the point of content capture. Although a user may design a story using existing content, a user typically designs part or all of a story first and then captures the appropriate content. By better understanding story structure and a user's intended message, computational tools may be able to provide effective advice or assistance at the point of capture. Giving users the ability to capture content is not sufficient; researchers much help them capture the right content in the right form for their story.
- Evaluating the use of multimedia in education, entertainment, instruction and training, storytelling, and more. A significant challenge for multimedia research is develop effective applications and validate their efficacy through empirical comparisons to traditional methods of delivering content. By providing more evidence that shows increased value with the use of multimedia, we can encourage more people to use multimedia more often.
- Viewing multimedia more as input rather than just as output. The multimedia research community often views multimedia as an end product, e.g., multimedia applications. The community should also think of multimedia, however, as a form of input that can enhance user interaction with computing systems often referred to as multimodal interaction. Although papers on multimodal interaction have been published in our conference, we should encourage more of this work to be submitted in the future. I believe that multimodal interaction is an emerging area of research that could offer more opportunity for the tools and applications area in the Multimedia conference.

4. State of SIGMM and the ACM Multimedia Conference

The state of SIGMM and the Multimedia conference has been and continues to be very strong. From reviewing the annual reports, the Multimedia conference continues to report encouraging net profits, even in the midst of an economic downturn, and will soon be operating independent of other SIGs. Attendance to the conference remains steady at about 225 attendees per year, however, it would be interesting to know how well the

conference retains newer attendees, which may be a more effective indicator of the strength of the conference.

5. Recommended Changes for SIGMM to Pursue

In the future, I recommend that SIGMM should continue to:

- Seek participation from young researchers in the field. This will help ensure that SIGMM conferences receive a steady stream of papers, reviewers, attendees, and volunteers in the future.
- Maintain a low acceptance rate for submissions to conferences, especially Multimedia, to help ensure that researchers perceive these conferences as first-submission conferences. By a first-submission conference, I mean that we want researchers to submit their best work to Multimedia first, rather than submitting revisions of papers previously declined from other conferences. A low acceptance rate helps researchers perceive Multimedia as "the place to publish" and should encourage them to submit their best papers to Multimedia first.
- Organize an ACM journal for multimedia. Although we have a journal, it is not a full ACM journal, because it is not listed or archived in the ACM digital library. An ACM journal would increase visibility of the field, satisfy peers who want to see ACM journals in publication lists, and offers a natural progression for research that was first published in the Multimedia conference.
- Lead by example. SIGMM needs a stronger web presence and should add more multimedia content to the SIGMM site. Researchers involved in SIGMM should challenge themselves to use more multimedia content in research and education and use their own tools as part of their research.

Dick C.A. Bulterman (CWI)

1. What are the fundamental topics/areas in the field of multimedia?

It is tempting to categorize multimedia research in terms of its established component areas: media encoding, media storage, media access, media transport, media rendering and overall presentation composition and control. My problem with this approach is that it blurs the added value of a distinct 'multimedia' field from each's component's individual research community.

Instead, I prefer to think of multimedia as considering the composite effects of creating and accessing and transporting and presenting rich media objects. The key 'multimedia' questions then naturally go beyond any of the 'simplex' questions relating to a component area. I used to feel strongly that mono-media and multimedia were decidedly separate fields of research. I've mellowed here, and now feel that multimedia research may well focus on handling a single type of media — such as images or even text — as long as the results can be scaled to time-sensitive media delivery or so long as more than one component area (encoding/storage/transport) are considered.

The following list gives a bottom-up collection of component topics and areas in multimedia.

- single/complex media format encoding
- single/complex media creation and editing tools
- single/complex media rendering engines
- single/complex media metadata description and semantic modelling
- single/complex media storage and access
- time sensitive network infrastructures and transfer protocols for single/complex media
- presentation description languages that de-couple presentation and media
- distributed control algorithms for selecting, accessing and presenting integration level composite media streams
- presentation description models (scripting vs. declarative, static vs. dynamic)
- presentation authoring and editing tools
- presentation verification and analysis (specification vs. runtime mapping)
- presentation metadata description and semantic modelling
- streaming end-to-end protocols for single/complex/presentation media
- presentation customizing (static and dynamic) for heterogeneous presentation platforms.

In this list, research into single/complex/presentation use of media can be differentiated. The presentation use of media differs from 'complex' media because presentations can inherently involve scatter/gather distributed access of media assets, while complex media objects typically contain multiple co-located assets.

2. What multimedia problems have been solved?

Single media creation, editing, storage and access is largely a mature area, supported by many commercial tools. There is plenty of room for engineering improvements of existing tools, but probably not a need for much new fundamental research.

The creation of glued composite streams is also pretty well understood (higher MP4 profiles)— where glued means a celluloid-like model in which the inter-media synchronization is guaranteed at authoring time.

Support for the streamed delivery of single or glued composite media objects across homogeneous networks is also pretty well understood (although implementations could improve). (The case for heterogeneous networks or heterogeneous users is anything but solved!)

Support for simplex hypermedia (that is, simple uni-directional links from single anchors) is also well understood, although the UI aspects for temporal and complex links are not.

3. What successes can research in this field claim?

If we compare the situation in 2003 with 1993, it is clear that the creation, transport and delivery of multimedia — specifically digital media — has become close to ubiquitous:

- *Single media*: a variety of formats have been adopted and deployed.
- *Composite media*: MPEG-4
- *Presentation media*: SMIL.
- *Network protocols*: RTP/RTSP.
- *Ubiquitous playback*: media renderers for single/composite/presentation objects on every desktop and (soon) on every telephone and PDA

Some of these successes grew out of fundamental research, others were guided and kept honest by research activity.

4. On what problems should researchers be focused?

While maturity is often accompanied by a 'the thrill is gone' feeling, there are plenty of interesting problems left to be solved. These cover all areas of media creation, processing and delivery. In my own area, the key issues are:

- Creating composite media and presentation descriptions:
- Support for adaptive presentations:
- Accessing information rather than selecting bits: navigation through complex spaces,
- Adding value to media without violating copyright and ownership:
- Integrating non-linear access into discrete and continuous media.

While each sub-area can probably compile a similar list, I also feel that there is an important are of research integration that also needs to be addressed. I would argue strongly that a common research delivery environment needs to be created as a reference platform for developing and evaluating new research. One of the reasons that MM research has limited visibility and impact (beyond that obtained within the research agendas of the component technologies) is that results are difficult to duplicate and

difficult to integrate. As a result, islands of competence are built, but individual research results rarely get expanded upon by other groups in a multimedia context.

(We have started on a project to create the basis for such an environment: a GPL opensource SMIL player environment that is free from commercial encumbrances and which can be used as a common platform for UI, codec, protocol and distributed access research.)

The last 10 years has allowed the media-enabled world to get access to one-size-fitsall presentations. The challenge for the next 10 years is to provide added value to media accessing and viewing in heterogeneous environments, so that the 'standard' media experience can be expanded beyond a shadow of what is available in the analog media world.

5. What is the state of SIGMM?

I have been out of the SIGMM loop lately, but while I regret some of the pending board departures, there seems to be sufficient continuity and energy for the future.

6. What is the state of the annual ACM Multimedia Conference?

The last conference I actively organized was in 1999. Since then, the numbers seem to remain very positive and the quality is definitely high. In spite of a wealth of alternatives (some of which are typically more festive and seem to be better funded), my impression is the ACM Multimedia is still the place to present.

7. What changes might SIGMM pursue?

I think it would be helpful if SIGMM would be more pro-active in establishing and pushing a research agenda to focus attention on important problems and themes. The specification of a set of grand themes (that get attention in the program of the yearly conference and get ink in affiliated publications) could be very useful to energize the MM research space. A series of small-scale workshops (not unlike the original NOSSDAV workshops) could be organized around these themes to build contacts and feed the yearly big conference.

In order to support the workshop model, it may also be useful to engage a set of regional activities (and even vice-chairs) for North/South America, Europe and North/South Asia.

Shih-Fu Chang (Columbia University)

What makes Multimedia a distinctive research field?

After thinking over this question several times again, I found the answer is still very simple – *Multimedia Research* addresses unique technical challenges arising from the joint presence and complex interaction among multiple media, including image, video, audio, speech, text, graphics, and other associated data.

The fact that there are synchronous or asynchronous data of multiple media brings about new challenges and unique opportunities that conventional fields have not yet or are not able to fully address. For example, in content analysis, a recognized emerging trend is the pursuit of multi-modal analysis and modeling. There is a tremendous amount of knowledge in individual fields such as computer vision, image processing, speech/acoustic recognition, and language processing, but how we harness different modalities effectively in understanding multimedia content is still a territory full of opportunities but yet not fully explored.

Some recent tasks clearly validate the above observation. In the recent TRECVID news video story segmentation task, approaches using multi-modal fusion clearly outperform those using single modalities only. In consumer video authoring applications such as automatic movie maker, a key requirement for optimal perceptual quality is to exploit the joint aesthetic characteristics of the visual and audio content. In producing the highlight and skimmed summary of video, joint consideration of audio-visual syntax and perceptual quality is a fundamental requirement. Solutions meeting such requirements cannot be drawn from individual conventional fields – a new breed of researches drawing on theories and techniques culled from different disciplines are needed.

Success stories in multimedia research exist but do not abound. Some multimedia representation, coding, and presentation technologies like DVD, SMIL, and Flash have enjoyed relatively broad adoption. On the other hand, streaming media, media content management, and media authoring technologies have not really taken off yet. The delay of success is not purely due to lack of mature technologies. Often, there are complex business and economic factors contributing to the delay. In the meantime, however, what should be the strategic directions of the multimedia research community in order to make multimedia a more recognized and sustainable field?

In my view, there are four critical strategic directions we should take to help establish a stronger and more recognized multimedia field:

1. *Focus on Critical Problems That Truly Involve Multiple Media*
   The community as a whole will receive increased recognition and credit if we can identify and focus on important problems that really require new solutions addressing multiple media. Breakthroughs in single media research (e.g., video coding, security and many others) will certainly facilitate progress in multimedia but do not directly contribute to recognition of the multimedia field. If we take a top-down view and consider practical applications, there exist many potential grand challenges for multimedia in various domains such as consumer, enterprise, and education. For instance, developments of consumer content management systems clearly require joint

analysis, exploitation, and manipulation of multimedia content, including photos, video, music, data, and documents. Multimedia meeting, training, and communication technologies and systems are also critical for enterprise and education applications.

2. *Establish New Theories and Algorithms of Multimedia*
   Multimedia applications and systems are often built by "borrowing" and/or combining existing theories or algorithms from existing fields. For example, many statistical modeling and learning methods from speech recognition and machine learning have been applied (or extended) to understanding multimedia content with various successes. Quality of service issues for video streaming have been tackled by applying novel algorithms from networking, operation systems, and databases. However, so far we have not been able to identify a large body of new theories or algorithms that can be clearly attributed to the multimedia field. As multimedia researchers, perhaps we should try to analyze the practical tasks at hands and make increased efforts to define, abstract, and formulate problems that truly call for multimedia solutions. At the same time, we should also identify and articulate what feedback and new insights the multimedia problems can introduce back to the contributing fields and facilitate their further advancement.

3. *Establish Rigorous Benchmarks and Evaluation Methodologies*
   Every recognized field needs rigorous quantifiable benchmarks and evaluation methods to assess performance, compare competing approaches, and identify new opportunities. Excellent examples include face recognition, information retrieval, and speech recognition. Unfortunately, so far the multimedia community has not been successful in establishing commonly accepted dataset, metrics, and benchmarking procedures. It is encouraging to see some recent efforts that have been taken, such as the NIST TRECVID benchmarking event for video retrieval. But much more of such efforts at a broader base should be made in all major areas of multimedia.

4. *Share and Accumulate Open-Source Tools and Knowledge*
   Open-source tool and knowledge sharing is key to the sustained growth of a new field. We have seen many successful examples in other fields such as Linux, speech recognition, and recently machine learning. In multimedia, there have been some respectable examples such as Berkeley's multimedia toolkit and IBM's video annotation and MPEG-7 description tools – but much more are needed. For example, one major barrier for new researchers to enter the media content recognition area is the steep learning curve and extensive implementation efforts required for feature extraction from multimedia content. Provision of an open-source library for such purposes will be of great value.

The above list is by no means exhaustive– there are many other important directions to be considered. For example, multimedia plays an increasingly important role in new generations of art, performance, and entertainment. Cross-disciplinary interaction between multimedia and these fields will provide fruitful paths of research and education. For another example, with the growing focus on the user-centric paradigm, incorporation of intelligence into multimedia systems that can be customized and adaptive to personal contextual needs is critical. However, taking the risk of being provocative, I consider the four directions presented above the most fundamental and urgent ones that call for immediate collaborative responses from the whole community.

Tat-Seng Chua (National University of Singapore)

The Use of All Available Knowledge in Developing Robust Scalable Multimedia Technologies

For many years, we have been obsessed with systems that understand natural human languages and able to engage in conversation with general users on any topics. This, however, has been proven to be too difficult. Instead, we found out in early 70s that simple term-based information retrieval (IR) technology works well and is able to scale-up to deal with voluminous texts. The IR technique, though works well in general search engines, has been found to be insufficiently accurate for many applications. To address this problem, information extraction (IE) techniques have been developed to extract semantic units of interests in specific domain. In addition, to ensure such techniques scale-up to wide range of applications, XML has been evolved to encode domain knowledge in a meta-level model. XML acts as mid-level representation that facilitates the automatic extraction of information from text documents using IE techniques and access of information at high levels. In addition to using meta-level domain knowledge, ontology, linguistic resources, and external and "redundant" knowledge from the web are frequently used to enhance information extraction accuracy, and to support "information" (rather than document) retrieval in question-answering.

The same picture can be painted for multimedia information retrieval and semantic information extraction. We have invested vast efforts in computer vision for many years with little success in scaling up the techniques to handle large dataset. Content-based retrieval brought some relief to the field in its ability to manage large volume of general image /video datasets. As with information retrieval, we soon realized that we are unable to improve the effectiveness of content-based retrieval techniques, especially those that based only on low level visual content features such as color histogram and texture. We were slow in introducing other content features for video such as audio and ASR (automatic speech recognition). Only recently we began to use these multimedia features routinely to analyze video contents. The use of only intra-content features, however, is still inadequate. To progress further, it seems that we need an XML-like meta-level model to encode domain knowledge, and we need judicious use of external knowledge, like the redundancy of web, ontology, and linguistic resources (dictionaries, encyclopedia) etc. The use of truly multi-media information and meta-level models has been advocated by Ramesh Jain recently.

To illustrate my points, let's consider the research topics of news and sports video analysis and retrieval. For news video, the important tasks that most researchers are tackling are semantic class identification, story segmentation and classification, and (personalized) retrieval. Because of commercial interests, various meta-level models for news have been developed, including the TV-Anytime standard. The TV-Anytime standard bases its low and mid-level content encoding on MPEG-7, but focuses on encoding the domain knowledge in terms of classes/events for news stories, and user access sessions. A rather comprehensive set of class/event taxonomy for news, commercial and sports have been defined. What is lacking is the modeling of contents within each class and relationship between classes, and how these classes may be detected and segmented. As in text processing, such enhanced model may be used as

meta-level model to facilitate the semantic extraction of contents defined within the model, and to support high-level access to video. It is clear that a range of multi-modal information, along with ASR, is needed to support such tasks. In addition, for news video, it will be effective to use external knowledge from the news web sites, and ontologies on names of locations and news-worthy persons. Recent research in CMU and NUS has demonstrated the importance of using such external knowledge to support the process of video text extraction, naming faces in news stories, correct ASR errors and precise video retrieval. Without using the full range of both intra and external knowledge, we are often left with insufficient information to analyze the video contents.

Recent activities in video-TREC (Text REtrieval Conference) focus on news video. It makes available 120 hours of news video and defines the tasks of semantic class identification, news story segmentation and classification, and retrieval. Video-TREC encourages the use of multi-modal data, including ASR, by also making such information available to all participants. It is a good beginning and provides a good basis for long-term research with ground truth for vigorous and large scale evaluations of robust multimedia technologies.

A similar meta-level model for domain knowledge can be developed for sports. For example, using the TV-Anytime standard as the base, we can extract detailed class/event taxonomy for sports. We can also incorporate event detection models, either explicitly, implicitly through learning approaches, or combination of both. A multi-modal fusion model can then be developed that uses the generalized domain model as the guide to extract mid-level features and events. In addition, for almost all sports that are of interests to users, there is post-match information widely available in news web sites that provide al least the timings and scorers of goals, in the case of soccer for example. Moreover, for important matches, such as the British premier league soccer, we can find detailed minute-by-minute commentaries on major sports web sites such as the ESPN. It provides detailed event commentary in text such as "Player X intercepts the cross of Player Y", and "Play Z scores a goal at xx minutes of the match" etc. Such external knowledge, at different level of details, may be used to support event extraction in sports video at different level of granularity. A scalable model can be developed that uses different levels of external knowledge, in additional to intra multi-modal content analysis, to extract events in sports video.

Similar scenario may be derived for other domains such as meeting, seminar and distance education.

Based on the above analysis, the important issues in multimedia in general, and content analysis and retrieval in particular, are:

- The fusion of multi-modal content features. It should emphasize the integrated handling of multi-media data, but not just analyzing uni-media data first, and performing fusion as an after thought.
- The use of higher order statistics (such as graphical model) to performing feature selection and fusion of multi-modal contents.
- The judicious use of external knowledge.

- The development of mid-level XML-like meta-level model to encode domain knowledge to guide semantic content extraction and retrieval.
- The research into robust multimedia techniques that tackles all stages (including pre-production, production and post-production) of multimedia content processing, and able to scale-up to large datasets with wide variety of domains. We should reduce obsession with fully automated techniques, and remove mis-conception that multimedia means visual media.
- The availability of large video data sets with ground truth to guide research into robust techniques and systems for wide range of semantic object extraction, and retrieval tasks. As video-TREC focuses on news video because of DARPA interests, our community could look into providing large scale ground-truth for sports, seminar, distance education etc.

In short, we should encourage research into robust techniques that use all available knowledge sources and are able to scale up to tackle large-scale problems. Only if we focus on these real-life issues would multimedia community be considered seriously as a field, like text processing, rather than a second rate cousins of computer vision, information retrieval and computer human interaction fields.

State of SIGMM and the ACM Multimedia Conference

On the SIGMM and ACM Multimedia Conference issues, we need to consider:

- The setting up of core under-graduate and graduate courses and curricula on multimedia.
- Move towards blind review of ACM multimedia conference with experience program committee members.
- Promotion of truly multimedia researchers in ACM and IEEE communities.

Nevenka Dimitrova (Philips Research)

1. Your view of the multimedia field - what are the fundamental topics/areas?

Multimedia field has been in an ever-evolving self-defining process. In the beginning it was anything but text: images, video, audio, and recently we started shifting to a definition of "multimedia" as something with multiple modalities. Bred from the parent fields of networking, systems, computer graphics, databases, signal processing and pattern recognition, the field had to grow into a separate own entity with own problems and solutions. Nevertheless, the topics seem a lot like the parent names with a multimedia prefix: multimedia signal processing, content-based information retrieval and pattern discovery, integration of media, art and multimedia technology, multimedia communications and networking, multimedia security and content protection, multimedia standards, multimedia databases and digital libraries, multimedia computing systems and appliances.

With the advancement of sensor technology any signal generated from different sensors in a variety of environments (e.g. user's home environment, medical, GIS, space) is added to the "media" basket of multimedia.

2. What multimedia problems have been solved? What are the successes?

It is sometimes hard to demarcate the commercial success from the impact of the research in the "multimedia community" to the commercial development. The widest impact has been made in wireless multimedia as a culmination of networking, coding, systems aspects: We have mobile phones with cameras! On the tools and applications side, multimedia collaborative applications (e.g. NetMeeting, Yahoo Messenger) are part of our business and private lives now. On the multimedia storage and retrieval side: Although we predicted that image retrieval will be most relevant, it turned out that audio content retrieval is becoming a huge problem right now with the available devices such as Streamium and iPod and services such as Rhapsody online music (paid) and Napster like services (free). In the near future we'll see the impact in the image retrieval and video retrieval with the availability of consumer multimedia PCs and similar devices.

In the future, the practice will have to include wide applications that use this technology in support of "normal" activities of the users: their everyday life, work and entertainment. In all three categories they could be served by storage, communications and productivity multimedia technologies.

3. On what problems should researchers be focused?

As a researcher in the area of multimedia content analysis and retrieval I will elaborate on the problems in this area. The trend in multimedia content processing, analysis and retrieval research so far was to use existing methods in computer vision, audio analysis and databases and make small steps in formalizing these solutions into solid mathematical frameworks. I want to divide the area into content analysis, feature extraction, representation and indexing, and potential applications.

Multimodal content understanding is primed to solve the holy grail of extracting semantics from features. Out of the above four areas this is the least explored area right

now. We have a fair understanding of computer vision technology its applications and limitations. We have also a fair understanding of speech recognition, but to a much less extent of audio scene content analysis and understanding. However, we still have rudimentary approaches to a holistic understanding of video based on audio, visual and text analysis and fusion of information.

In the future, multimedia research will have to break new frontiers and extend the parent fields exactly in the area of multimodal processing and information fusion. To achieve this, we need to rely on context and memory. Context is the larger environmental knowledge that includes the laws of biology and physics and common sense. In philosophical terms so far, we have been using what I call the "Hume" model of signal processing where the only things that exist in the present frame are real, and we should transcend to the "Kant" model where there is a representation which accounts for common sense knowledge and assumptions about the expected behavior of the entities that are sought for. Memory is important aiding factor in analysis with longer term goals. In this respect our methods have severe anterograde amnesia and we just keep a very localized information about the current computations. In detection of dissolves, we keep a buffer of frames. In computing scenes we keep a few minutes worth of data. However, in multimedia processing we need to keep more information for longer periods of time, such as full programs, episodes and genres.

We need to consolidate the theoretical foundations of multimedia research so that it can be considered on an equal footing with the parenting fields. We need to go beyond the simple application of pattern recognition to all the features that we can extract in compressed and uncompressed domain. We need new pattern recognition techniques that will take into account context and memory.

4. What is the state of SIGMM and the ACM Multimedia Conference?

Currently SIGMM is a small and not really active entity and there is no incentive for people to join the group. ACM MM is the recognized prestigious conference on multimedia with a rate of about 200 attendees every year. The conference itself has developed its own community of people working on the established topics of content processing, applications and systems.

5. What changes might SIGMM pursue?

The SIGMM has to become more lively and desirable for people to join. We have to create an active community of people that will draw on the community experience of other people in the SIGMM. It is very important to draw in researchers as well as practitioners just like SIGGRAPH does.

The ACM Multimedia conference needs to expand the scope with new topics related to multimedia and to increase attendance.

Wolfgang Effelsberg (U. Mannheim)

1. My view of the multimedia field - what are the fundamental topics/areas?

My definition of a multimedia system: a system that integrates discrete media (such as text, still images) and continuous media (audio, video, animations). Fundamental topics are:

- compression of discrete and continuous MM elements, in particular audio and video streams
- content analysis of MM elements (in particular video), i.e., trying to understand the semantic contents of MM data steams. Includes automatic indexing, information retrieval for unformatted data objects, automatic abstracting, etc.
- multimedia communication, i.e., the transmission of MM data streams over digital networks
- multimedia file servers, in particular video servers
- data structures for MM presentations (such as SMIL)
- MM applications, such as tools for MM authoring, e-learning with MM documents, voice over IP, videoconferencing tools
- and more

2. What multimedia problems have been solved? What are the successes?

- The integration of audio into every computer
- Audio compression that allows everyone on the Internet to retrieve HiFi stereo quality audio files
- Video compression with ratios of 100:1 or better, at a high level of quality
- Still image compression with ratios of 100:1 or better, at a high level of quality
- Multimedia elements on Web pages
- Powerful video file servers
- Cut detection

3. What multimedia problems have NOT been solved in spite of major efforts?

- QoS for MM streams on the Internet
- Multicast for MM streams on the Internet
- Fully automated analysis and indexing of visual content. Querying of MM databases by visual/audible examples
- Replacing traditional TV distribution channels by multi-purpose broadband networks that also transmit live TV streams
- Video-on-demand as a widely deployed service

4. *On what problems should researchers focus?*

- MM on mobile/wireless devices, in particular smart phones and PDAs
- MM user interfaces. We still lack a convincing metaphor to communicate with a MM computer; the desktop metaphor is obviously inappropriate. In particular, get rid of the keyboard!
- Content repurposing for different end systems / form factors
- Even better audio compression and voice compression
- Even better video compression. For example, wavelet-based compression was discovered relatively late and works quite well.
- Extracting semantics from video FOR SPECIFIC APPLICATIONS (e.g., traffic, face recognition, gesture recognition)
- Model-based face animation, in particular to show emotions; personalized avatars
- Virtual presence for videoconferencing
- Augmented reality
- MM art

5. *What is the state of SIGMM and the ACM Multimedia Conference?*

SIGMM

- A very active bunch of dedicated people, a lot of fun to work with
- Running a great conference and three very good journals/magazines with IEEE, ACM/Springer and Kluwer. A strong community, but in need of fresh blood and truly innovative ideas

The ACM Multimedia Conference

- Very strong from a scientific point of view, the best in the field worldwide. Much referenced
- Content-wise: stagnation, a lack of really exciting new ideas
- Industry involvement has never been very strong (unlike in SIGGRAPH or SIGCOMM)
- Shows signs of a mature field

What changes might SIGMM pursue?

- Try to increase industry involvement
- Try to increase representation of East Asia, at all levels of SIGMM. Run the conference in Singapore or Hong Kong or Shanghai or Taipei
- Try to include games as a topic
- Try to run a series of more focused, smaller workshops throughout the year, throughout the world

Jim Gemmell (Microsoft Research)

1. *Your view of the multimedia field - what are the fundamental topics/areas?*

The term "Multimedia" has usually meant IO involving more than just text (images, audio, video, animations) - but not computer graphics. Ideally it should be multi-media, i.e. manage or fuse multiple media types. But, aside from video that trivially includes an audio stream, this is usually not the case.

Multimedia is an applied field. It is more descriptive to call it multimedia systems and applications. It applies other fields to building systems and applications that support multimedia. These fields include: networking, OS, signal processing, content analysis (e.g. computer vision), information retrieval, hypertext, computer graphics, authoring and user interfaces.

The problem for our community is to define what constitutes interesting application/system, and whether a work is really not at home in some other field (especially computer vision or signal processing). We do not want to publish a paper that was just not good enough for publication in its field. We need some criteria by which to decide whether something is a multimedia work rather than just something belonging in a more specific field. For example, we might ask:

- Is the application broad enough, i.e., across enough disciplines, that no single discipline will publish it?
- Is it truly multi-media?
- Is it new media or media that we don't currently support well?
- Is the multimedia application just a thin wrapper over the true focus of the work (some technique that is field-specific), or is it really what the work is all about?

2. *What multimedia problems have been solved? What are the successes?*

O/S & network support for delay-sensitive media (primarily audio & video). Capture, storage and streaming of video are now mainstream consumer applications. We can look back on contributions in authoring (SMIL), scheduling, network protocols, and adaptive streaming formats.

3. *On what problems should researchers be focused?*

I'll re-word this as: what problems have not been solved? What are the failures?

- Useful videoconferencing.
- Low-latency, high-quality audio communication using the Internet.
- Personal media management
- Authoring
- Support for video is still not on par with photos, and certainly not with text.
- Scaling in a number of dimensions: many capture devices, sharing among many individuals, etc

- We can also explore enhanced or different kinds of media like stereo images/video, more aspects of spatial sound, large (and/or curved) displays, and stimulating other senses (smell, touch).
- Hypermedia systems and authoring

4. *What is the state of SIGMM and the ACM Multimedia Conference?*

I think the decline in numbers is partly due to what is already seen as solved, and as part of the declining value of membership with the communication that happens via the Internet. The content is still strong and interesting, but sometimes includes work that really belongs in a more specific field.

5. *What changes might SIGMM pursue?*

We need to publish guidelines by which we will consider a paper to be suitable for SIGMM, in contrast with a paper from some other discipline that is commonly applied to multimedia. These guidelines should be updated each year or two. Such guidelines happen now inside program committees where the decision making is not transparent. Furthermore, authors only find out what the rules of the game were after their paper is accepted or rejected.

It is an embarrassment that the publications for the multimedia community are on paper. Both the conference and the journal must be moved to the web (with video, audio, SMIL supported). A deal should be made with someone like Kinkos to print/bind/ship on-demand. The conference talks and SIGMM meetings should be video recorded and also published on the web. Live streaming would be good publicity, but is probably not really that important

Nicolas Georganas (U of Ottawa)

1. *Your view of the multimedia field what are the fundamental topics/areas?*
   - The field is getting very mature in the traditional areas of audio and video;
   - New media have been added to the traditional ones. In addition to images, text, video, audio, graphics, animation, we now also have Virtual Reality simulations, haptics, sensor data fusion, Kansai communications (at the ATR International labs in Kyoto), smell, holograms, and who knows what other exciting things. I believe that we are fast approaching the creation of Ray Kurtzweil's "Sensorium", i.e., a multimedia kiosk with total senses' immersion.
   - Multimedia interfaces remain a very exciting area for research. Context-aware speech and gesture recognition, and Brain-wave/thought recognition will be most exciting interfacing areas;
   - Interactive and collaborative multimedia data distribution over networks with poor QoS will remain a challenge, with the onus on the application protocols to do latency compensation and/or prediction;
   - Digital Watermarking will remain a prominent area for research, as every new system is limited in its power and challenged by clever attacks.
2. *What multimedia problems have been solved? What are the successes?*
   - I think that the issues pertaining to Multimedia Synchronization have been solved and the topic exhausted. Good algorithms exist and are sufficient.
3. *On what problems should research be focused?*
   - See items in 1 above.
   - Multimedia data sensor fusion and sensor networks.
   - Image-based searches
4. *What is the state of SIGMM and the ACM Multimedia Conference?*
   - ACM Multimedia Conference: Quality very high, selection process strict but attendance low. Some mechanisms for bring back the early successes in San Francisco with 1,000 participants.
   - SIG MM: Basically running the Multimedia conference and sponsoring some other affiliated conferences. We should brainstorm its future mission and goals.
5. *What changes might SIGMM pursue?*
   - (No suggestions at this point. I wait for brainstorming).

Forouzan Golshani (Arizona State University)

1. Your view of the multimedia field - what are the fundamental topics/areas?

Clearly at the heart of the multimedia field are media processing and authoring tools. Along with these we require the capability of real-time content analysis, indexing, and integration, which necessitates cross-modal ontologies. Having these, we can then emphasize personalization of multimedia information, with the associated problems such as content protection and security.

2. What multimedia problems have been solved? What are the successes?

Media streaming has reached a pretty good degree of sophistication. Also we have seen significant progress in retrieval capabilities. From application point of view, we can site novel applications in areas such as education, caring for individuals with disabilities, and the arts (beside the obvious one, namely entertainment.)

3. On what problems should researchers be focussed?

Media content analysis and representation (and all of those good things that MPEG7 talks about), Interactivity with media objects, and, probably more important than anything else, commercialization of so many neat ideas that our colleagues in this field have generated.

4. What is the state of SIGMM and the ACM Multimedia Conference?

ACM Multimedia is considered to be the premier conference in this field. Unfortunately, I have not been so active in the SIGMM, so I cannot comment on it.

5. What changes might SIGMM pursue?

Allow sponsorship of timely workshops and conferences.

Martin Kienzle (IBM Research)

1. What are the fundamental topics/areas in the field of multimedia?

The objective of multimedia technology is to integrate media intuitively and easily into digital systems, and to use the unique benefits of media to make existing media uses more satisfying and productive, and to enable and invent new uses of media.

The major fields in media research are: media representation, processing & compression, creation & composition, analysis, management, search and retrieval, distribution, middleware, and applications.

2. What multimedia problems have been solved?

The key to the answer of this question is the definition of "solved". I am defining "solved" as "used widely in recreational, entertainment, educational, or commercial applications."

Under this definition, only a small subset of multimedia problems have been solved: media representation, media processing, compression / decompression, streaming servers and protocols, media players, and editors (low-end as well as commercial grade). These are all "low level" component technologies. While there is a large body of knowledge and technology on many other topics, such as media analysis, media databases, search and retrieval, and integrated media applications, not much of this technology has made it into the main stream.

3. What successes can research in this field claim?

Major research successes (again, defining "success" as wide use) in multimedia are media processing, compression / decompression, and protocols. Most of the other "solved" problems have been solved simply by commercial development, for instance in the areas of media editors and media players.

4. On what problems should researchers be focused?

Multimedia systems typically have far more moving parts than most other computer based systems. Most research is focused on one of the moving parts, with little consideration for the others. This creates point technologies that are difficult to assemble into complete systems. However, only complete systems will bring value to the users. Where researchers are building complete prototypes, it is still difficult to use these prototypes in situations other than they ones they were designed for. This leaves much outstanding research technology "stranded" due to lack of a context.

To move forward, it will be essential to develop a high-level systems architecture, and to build a body of re-usable parts that support this architecture. On this foundation, researchers can build new systems, proving their unique new technologies in the context of complete, operable application systems. This approach also requires an accepted set of "glue", a collection of middleware, APIs and protocols, to connect the components of our systems. Finally, outside entertainment, media receives its value from its application context, such as teaching systems, surveillance systems, or business collaboration systems. We need programming technologies and middleware that allow us to create

systems that integrate media function seamlessly with general application function. This will help to create acceptance for our technology, and value for our users.

In order to make use of the large body of excellent science and technology that has been created we need a substantial dose of engineering to put it all together. This will make it easier to create the complete systems needed support usable applications. Operable applications will allow us to evaluate our research advances against the needs of users, rather than against other research results. We will be able to demonstrate the cost effectiveness of our systems (not just of our components!), and we will learn what aspects of a problem we may not have thought of, such as ease of use to create a large user base.

5. What is the state of SIGMM?

SIGMM is healthy as a body that supports research in multimedia, aids the exchange of research ideas, and promotes excellence in individual areas of research. However, it is not clear that SIGMM can maintain its viability as a separate discipline without more impact on the real world.

6. What is the state of the annual ACM Multimedia Conference?

The ACM Multimedia Conference is the premier conference for research in media technologies. Similar to SIGMM, the Conference is somewhat isolated from the commercial reality.

7. What changes might SIGMM pursue?

Some the changes to consider are:

- Start a discussions / interest groups about the engineering problems that need to be solved to bring the science results to use. This includes middleware, protocols, APIs, and a common high-level architecture. Create an engineering track for ACM Multimedia.
- Create a repository of components that researchers can use to assemble systems. I realize that some beginnings of this exist, but it can be done more systematically. (Just think back to the incredible explosion of usable technologies that the early spread of UNIX sparked!)
- Create "interop" events where researchers come together to demonstrate how their systems work together. These events could also be combined with competitions in terms of efficiency and performance. The Video TREC competition sponsored by NIST could provide an interesting model.
- Give recognition to interoperable research systems, such as an award at ACM Multimedia.
- Create a "mixed technology" track that explicitly deals with integrating media technologies with "standard" application technologies.
- Vigorously support the exploration of new application domains. The work on sensors, surveillance, and low-power devices are good examples of this type of work.

Joseph Konstan (U of Minnesota)

My research is centered in human-computer interaction, though I have conducted MM research (including publishing in ACM Multimedia) when conducting research that overlaps the community (command streams, multimedia authoring tools).

My service in this community has included serving as one of the Program Chairs for ACM Multimedia 2000, running the first Dissertation Symposium at ACM Multimedia '99, serving on the SIGMM executive committee since 1999, serving on the board of the Multimedia Systems journal since 2001, and various other positions in conferences. I am also active outside SIGMM, including currently chairing both ACM SIGCHI and the UIST 2003 conference.

My bias, which will become clear soon, is that multimedia as a field of computer science cannot survive unless it embraces important technical and scientific challenges that are unique to the composition of multiple media types. While there have been, and still are, challenges in image processing, networking support for video, and other media handling and transmission, these challenges do not bring together people -- rather they are better addressed in communities that deal with image processing and computer vision, networking, etc. On the other hand, challenges such as authoring, cross-media adaptation, and even cross-media delivery and service levels, make sense as the core of Multimedia research.

1. My View of the Multimedia Field

Multimedia Computing (which is really what I think our field is) is about the unique challenges in computer science and its applications that arise when attempting to provide rich interactions that involve the coordinated use of multiple media. In some ways, this places Multimedia Computing outside the taxonomy of "field" and more into the taxonomy of "cross-disciplinary domain." Some examples of things I see as fundamental are:

- Use of multiple media for indexing and retrieval (e.g., using video, images, text, and audio together to index into video libraries)
- Authoring of multimedia -- this seems fundamentally hard once it gets beyond very simply capture
- Interactive multimedia -- the technology, tools, and interfaces for delivery of interactive multimedia experiences, including those in gaming, education, and simulation environments
- Cross-media technology issues -- I think some of the issues in file systems, networking, etc., are really issues that relate primarily to the challenge of having multiple media (either for dealing with levels of service or for common access patterns).
- Human factors of multimedia -- research questions on the effectiveness of multimedia for specific purposes, human perception of multimedia, etc.

2. *What Has Been Solved*

- Video-in-a-window (even with audio) -- this is a long-ago solved problem that is now a clear historical success.
- Video compression, and video manipulation in the compressed domain -- this is impressive work, but it is now post-research and another success.
- Basic hypermedia specification -- SMIL and a host of research systems have shown the ability to specify multimedia and hypermedia.
- Cross-media indexing and search -- while there is more to do, there is also a great deal of success in this area, showing the value of mixing captions with images or audio with video.
- Video servers -- we did it. Now it is a commercial solution.
- Quality-of-Service and networking for multimedia -- great work, now we can buy it.

3. *Where should researchers focus?*

- Rich and interactive media -- there is still a world beyond audio and video, and few are exploring the media of interactive spaces, or even the potential interactions with existing media.
- Balancing automation with manual control -- many systems (e.g., editing, conferencing, classroom capture) either provide a high-effort operator mode or a low-flexibility automated mode. A fertile area for research is the design of systems that provide guidable automation, so that an operator (or user) can "steer" the system in the right direction, without being trained in or consumed by all the technical options.
- More user-validation of research. Too many studies in this field neglect the question of "does it actually work for people." This includes the question of whether a measurable difference is perceivable, but also whether classification and query systems are good enough for people to use.
- More complex application-driven research. Just as the challenge of "Eye Vision" at the superbowl led to advances in robotic camera operation, and as challenges in terrorist-detection have pushed real-time face recognition forward, we need to take on challenging applications to really push multimedia technology to its limits.

4. *What is the state of SIGMM and the ACM Multimedia Conference*

Healthy, but precarious. Collectively, the SIG and the Conference have the following serious challenges to address:

- Much of the best work "in the field" happens outside of this community. People doing multimedia data fusion (e.g., Mitre's work on gleaning intelligence from television news), and others doing interesting work (including the "eye vision") don't necessarily see a multimedia conference or group as their home. The same is true for the large amount of multimedia-related HCI work (which often gets

published at CHI), and IR work (which often gets published at SIGIR). To be fair, there is still a lot of good work submitted to and published at ACM Multimedia, but that position isn't solid across the domains of multimedia research.

- ACM Multimedia isn't the primary conference for enough of its attendees and authors. As budgets are cut, there is a significant risk that people will skip ACM Multimedia for a conference on machine vision or pattern recognition or HCI or Networking or ...
- ACM Multimedia still has to contend with competing conferences in the same field. While I think it is by far the highest-quality conference, it is unclear whether that always guarantees success when competing with other conferences.
- It appears that SIGMM and ACM Multimedia are "irrelevant" to industry.
- Multimedia Systems is broken. ACM doesn't own the journal, and seems in many ways to have disavowed it (there are many stories, but one repeated theme is that they lost faith in it when it was too far behind in publication). Springer won't hand it over (or sell it). It is time to abandon it for a new ACM Transactions on Multimedia.
- The SIG, and the Conference, are more of a confederation than a melting pot. Based on my experiences in recent years (and not-so-recent ones), there is far too little cross-fertilization between the "applications and interfaces" people, the "systems" people, and the "content" people.
- On a related note, it is unclear how much of a cadre of "cross-over people" exists to lead the organization. It would be a terrible shame if SIGMM and the conference couldn't keep drawing together these communities, and part of that requires leadership that itself draws together the communities.

5. *Changes to Pursue*

- Create an ACM Transactions (and if possible, convince the leaders of the field to abandon MMSJ).
- Consider a periodic co-location with a conference with large industrial attendance (e.g., SIGGRAPH) to encourage building bridges back with industry.
- Find out more about who the membership really is, who they identify with, and what they want.
- Spend effort bringing together diverse fields. Create more conference sessions that cross tracks but elaborate on common themes.
- Consider having a newsletter that includes "reviews" of work published and presented elsewhere. "Multimedia tidbits" could help keep people listening to SIGMM when they're looking for the best work in the field.

Ketan Mayer-Patel (U of North Carolina)

1. The State of the Field

In order to foresee the future of multimedia, I think it is important to identify what motivated the formation of the sub field in the first place. In my opinion, the field has tackled problems in the intersection of more traditional sub fields (i.e., operating systems, networking, graphics, coding theory, vision, etc.) because its driving applications demanded integrated solutions in order to achieve desired results. In other words, the application was somehow more than just the sum of its parts.

Thus, when we experimented with the first Internet-based video conferencing applications over a decade ago, it was important to investigate and understand how video representations could be made network aware (and vice versa), the demands of media processing on the operating system, the relationship between video and audio from a perceptual quality point of view, and so on. The lessons learned developing systems that negotiated these complex tradeoffs are the real successes of multimedia as a field and where our contribution has been most fundamental.

In order to keep the field moving forward, we must seek out the applications of the future that demand integrated solutions with complex tradeoffs. Video conferencing is no longer a candidate. Nor is video-on-demand. While these applications once tested the limits of the systems we could build and inspired interesting new multimedia research, Moore's law, larger storage capacities, and fatter pipes have made these applications less compelling. A poor solution to these problems that inefficiently and naively throws resources at the problem will now work as well as a well-crafted, elegant, and innovative solution.

Given advances in processors, storage, and networking, then, what kinds of new applications now contain the types of challenges that video conferencing and video-on-demand once did? I believe there are a few characteristics to look for. First, we must be true to the prefix multi-. Multimedia of the future will not be video and audio, multimedia of the future will be tens of streams with complex interstream semantic relationships such as the thirty or so cameras used to capture almost every conceivable angle of the Superbowl. Second, we should broaden our notion of media. We need to include bioinformatic sensors, stock tickers, sports scores and statistics, and other sources of periodic, correlated, information in the same framework as video and audio. Third, interactivity and the user experience will remain a source of challenges. Fourth, rapid advances in graphics hardware, capture devices (e.g., cameras and microphones), and display devices (e.g., digital light projectors) should inspire us to build applications that employ them in new and different ways.

The following is a list of application areas that I believe have many of these characteristics and will drive the future of multimedia as a field. Clearly, this list is far from complete or comprehensive and reflects my own personal research biases:

- Large-scale peer-to-peer streaming
- Distributed virtual environments

- Ad hoc home-based media environments
- Teleimmersion
- Low bandwidth sensor-based applications
- Gaming
- Multimedia information retrieval and management systems

2. The State of the Conference

The cornerstone of our community of researchers has been and should continue to be the ACM Multimedia conference. We have a lot to be proud of. Acceptance rates are extremely competitive keeping quality reasonably high. The doctoral symposium has been a great success. The three track system has been useful for managing submissions from across such a wide array of areas.

In order to ensure its success, however, we must think creatively about how we organize and run this annual meeting and not be lulled into complacency. I describe three specific suggestions below as examples of kinds of changes that we might consider which I hope will spark discussion.

- Organize and run a challenge competition. The idea here is to define a specific challenge environment and inputs for a type of research problem. For example, scene detection or application-level multicast streaming may be target problem areas. As part of the call for participation, a specific API and metric used for success is published. Researchers are invited to participate by submitting software solutions (or better yet, the address of a web service interface to their solution). At the conference, the results of the challenge competition are presented along with an award to the winning system.
- Invite a panel of outside researchers in a related field. Each year, find three leaders in an area related to multimedia (vision, graphics, networking, operating systems, etc.) that have never attended ACM Multimedia to participate in a panel in order to give their view of how multimedia problems intersect with their field. Allocate a modest travel and accommodation budget and waive the conference registration as an incentive to participate.
- Invite on-line feedback and discussion. The presence of wireless connectivity at conferences is becoming increasingly commonplace. We can employ this resource to provide on-line chat rooms and instant messaging between conference participants. Transcripts of the public chat room can be made available so that presenters can review the discussion and reply later in the day. Anonymous feedback and discussion may be worthwhile feature.

Some of these suggestions are bit off-the-wall, but I think that creative experimentation is the key to keeping the conference fresh and interesting.

Klara Nahrstadt (U of Illinois Urbana Champaign)

1. My view of the multimedia field - what are the fundamental topics/areas?

The multimedia field includes currently nine major areas: (a) multimedia operating system; (b) multimedia networking and communication; (c) multimedia middleware and software engineering; (d) multimedia databases; (e) multimedia application services; (f) multimedia human computer interfaces; (g) multimedia coding and compression; (h) multimedia security; (i) multimedia technology in education and other collaborative environments.

2. What multimedia problems have been solved? What are the successes?

I will comment on multimedia systems and networking problems because I am working in this area. The following multimedia problems have been partially solved (none of the problems are fully solved as the computing and communication infrastructures are changing and existing algorithms must be augmented or replaced by new approaches): (a) understanding of multimedia synchronization; (b) understanding of multimedia retrieval, streaming and playback, (c) multimedia storage; (d) multimedia compression, (e) multimedia distribution through IP multicast and end-system multicast.

I consider the following multimedia results as great success: (a) multimedia storage (e.g. CD, DVD), (b) multimedia streaming protocols (e.g., RTP, VIC, VAT); (c) multimedia support through EDF scheduling, real-time extensions in current general purpose operating systems to support multimedia applications (e.g., multimedia support in Windows OS, Linux OS, Sun OS), (d) multicast algorithms driven by video conferencing, video on demand, and other multimedia services, (e) MPEG players (e.g., Berkeley MPEG player, Windows Media Player, MP3 players), (f) development of video and audio compression formats and standards (e.g., MPEG standard family, H.263, wavelet)

3. On what problems should researchers be focused?

The researchers should concentrate on the following problems: (a) ubiquitous and easy multimedia distribution infrastructure to allow for easy setup of video conference, video phone conversation, distributed lecture recording, and playback; (b) easy and automated mappings between multimedia applications and underlying system and networking infrastructure that would support and enforce the required quality of the multimedia services, (c) configurability and service management of multimedia applications over different system and networking infrastructures, (d) easy programming of complex and distributed multimedia applications, (e) energy-efficient and resource-constrained multimedia processing and communication for small hand-held devices and group of small devices (e.g., cell-phones, PDAs), (f) new types of multimedia applications (e.g., in digital homes, smart seminar/conference rooms, lecture rooms, medical domain, teleimmersion).

4. What is the state of SIGMM and the ACM Multimedia Conference?

The SIGMM provides strong support for the leading ACM Multimedia conference as well as for the ACM/Springer Multimedia Systems Journal. It is also supportive of

related multimedia conferences such as NOSSDAV and SPIE Multimedia Computing and Networking conferences. ACM Multimedia conference has very strong submission numbers, and it is the premier conference in the field. What might need to be improved is (a) the attendance (although after the dot-com boom/faculty returning from start-ups/ we may see increase in attendance), (b) the strong differentiation from conferences such as image processing, speech processing, DB, HCI conferences. I think, the differentiation happened at the systems and networking levels (to some degree), but at the application level, it is not clear if we are getting rejected papers from related conferences or not.

Attendance might be improved through inclusion of additional sessions/awards such as Work-in-Progress Sessions attracting students (this is in addition to the Doctorial Symposium). The Work-in-Progress session(s) could have 5-minute presentations for authors to get the main ideas across and get also feedback from the audience. We could also consider rewarding authors not only for best conference paper, but also best demo/poster, best work-in-progress idea, etc. We could also provide multiple awards for the best papers (e.g., three awards - best paper in applications, coding, systems). All these approaches might attract authors to submit as well as to attend the conference.

Strong differentiation could happen through a specification of topics in call for papers which would include emphasis on handling of two or more continuous/discrete media, or support for continuous traffic, or integration of two or more components in a multimedia system/application (e.g., handling of image processing and databases), or other clear differentiation that will not included in DB or HCI or Image Processing conferences.

Multimedia could be understood either handling two or more media, consider two or more pieces of the overall multimedia system (this could mean to look at single medium such as video, but integrating view of application and OS, application and network, etc).

5. What changes might SIGMM pursue?

It would be good to increase the visibility of SIGMM through being present not only at ACM Multimedia Conference, but also at other supported workshops/conferences (NOSSDAV, MMCN, HCI-related conferences, etc.) to give overview and inform about activities of SIGMM. This could be done through SIGMM lunches. SIGMM web page could provide more information about multimedia specific jobs, multimedia successes, multimedia books, multimedia activities, related conferences, workshops. Award(s) for achievements in multimedia would be of great motivation to the area and researchers in this area (e.g. SIGOPS gives out Mark Weiser award).

Arturo Pizano (Siemens Corporate Research, Princeton)

In 1995 the Siemens Corporate Research (SCR) established the *Multimedia & Video Technology Department* as a spin-off of the Imaging Department. This was an early recognition by our management that the then nascent field of Multimedia held significant promise as the source of innovation for the company's products services and solutions. As a point of reference, SCR is part of the Central Technology Department of Siemens AG, a global conglomerate with businesses in a wide range of fields including Medical Equipment and Services, Power Generation and Distribution, Transportation, Industrial Solutions and Telecommunications. The reminder of this paper describes the experiences we have gathered in our role as corporate technology drivers for multimedia.

As a technology promoter I now often find myself trying to define what we considered multimedia technology. This is in a stark contrast to earlier experiences when we had to explain what multimedia was. Perhaps this is the first point of reflection. In as much as we now associate multimedia technology with the ability to combine different types of media into a single presentation or object, in the strictest sense almost all the tools we now use are multimedia-enabled. Yet, we are often forced to subdivide the field into specific segments, whether it is to structure a conference, a publication or an R&D organization. In SCR we created a horizontal divider between infrastructure (including networking, encoding, streaming, security…) and applications, which is the focus of our work. For the latter we use a workflow-driven separation with four major programs: Acquisition and Authoring, Processing & Manipulation, Management & Distribution and Communications & Collaboration. While this is an arbitrary, and probably incomplete segmentation forced by organization and business directions it provides a useful framework for the remainder of the presentation.

Acquisition and Authoring

This is where the most tangible advances have been made. We have gone a long way since the acquisition of digital video required special boards and powerful workstations. Digital image and video cameras are now common even in mobile phones. Efficient encoding mechanisms have been widely adopted and incorporated into the software systems that we use to combine raw media into usable objects. Clearly, the multimedia field has been successful in this area; even if our ability to acquire media has exacerbated other problems, e.g., content management.

Processing and Manipulation

A great deal of effort has been placed on automatic content processing and semantic feature extraction - the field's Holy Grail. However, we have found limited success solving the general problem. Instead, we have learned that domain-specific knowledge plays a crucial role in enabling the efficient processing of multimedia content. Furthermore, there must be strong market motivation for pursuing automatic content processing in a particular field; this is because the efforts needed to properly address the problems are enormous.

At SCR this has clearly manifested in the development and evolution of the medical imaging domain. In this case the company has invested considerable resources over a

period of many years and the results are a key component of the company's products. By contrast, early efforts in generic automatic video processing (e.g. scene change detection) were comparatively minor and as of now do not play significant role in products.

Not coincidentally, topics in medical imaging/video have always been addressed in their own specialized conferences. Of course, the field was established long before multimedia, but we can also see they have reached critical mass in terms of the depth hand breath of the problems being addressed. A question that we could address in the workshop is how much more should the multimedia research community push for general solutions, or specialized solutions for which the business motivation is lacking? While it is always dangerous to inject business too early into the research process, the evidence shows that resources will not be available unless the community perceives a significant business opportunity. As anecdotal evidence for this position I can cite the growing interest in real-time video processing for security and surveillance at SCR, where a separate department was recently created.

Management and Distribution

As indicated above, the proliferation of acquisition devices and authoring tools has been a success story in multimedia. On the other hand, this has created challenges in the area of management and distribution. There are several areas where we have experienced success both within SCR as well as in the industry. In the area of management we have benefited from the integration the content authoring processes into the management systems that we use, and in the separation of internal content representation and presentation. By thinking ahead of time about management problems it is possible to create media objects that contain information useful for manipulation. Similarly, the ability to postpone decisions on presentation until the media is needed offers advantages when the target device is unknown, or where content repurposing content creates increased efficiency.

One of the success stories we have experienced at SCR is the multimedia documentation program, which specializes in addressing the particularly challenging issues involved in creating a digital representation of the enormous amount of information associated with a complex piece of equipment, such as a gas turbine. Without the ability to properly capture and author the basic content (manual, drawings, videos, 3D models…), and latter delivery it in the appropriate form (e.g., a CD-ROM, web-page or technical brochure) we would have hardly made a dent in the considerable inefficiencies that existed when paper was used.

This success notwithstanding, the challenges that remain are substantial in cases where we cannot control ahead of time the structure and presentation of content. The primary example is the World Wide Web, where the difficulty in finding the right information increases on a daily basis. While we are experiencing progress on text-based methods for indexing and retrieval, we are still far from being able to do content-based searches. This can be directly attributed to the problems we mentioned in automatic processing and feature extraction.

On the distribution side, the World Wide Web has also been a very strong driving force for the creation of efficient encoding and delivery mechanisms. Anticipating the availability of sufficient bandwidth the research community and industry have already developed robust streaming services. The problem has not been completely solved,

particularly as we anticipate new services will be deployed, e.g., true video-on-demand, but we have developed a solid understanding of the underlying problems. Cooperation with the networking community, and the identification of viable business models should motivate further research efforts.

Communications and Collaboration

The way in which we interact with each other using multimedia content has also advanced significantly during the past 8 years. Certainly the World Wide Web has been a catalyst for this progress, but multimedia applications such as e-mail and office tools have contributed significantly. Here too, the research and industrial community have been able to take advantage of advanced in hardware and software to offer more compelling applications. Even sufficient network bandwidth, which has taken longer to materialize, is now a reality. Long sought services, such as enabling remote document-based multimedia collaboration, are now commercially available.

Despite this progress opportunities still exist, particularly in the area of wireless communications. The advent of wireless networks, in particular WLAN, provides a reach platform on top of which we can realize applications already identified, but until now unfeasible. In addition, additional services such as presence and availability should motivate the development of completely new services. In this context our experience has been that application level research requires substantial links to the HMI community. It is easy to take a technology-driven approach that will result in what appears to be a very attractive solution, just to find out that a simple user factor, such as the unwillingness of may people to talk into a computer, will prevent the solution from being adopted, at least in the way it was originally envision. I have witnessed first hand how teenagers are much more likely to use text chat than voice chat. By working closely with our own User Interface Design Center we have been able to understand and anticipate this kind of reaction. Perhaps this is approach is too pragmatic and business oriented, but it would not hurt to periodically share experiences with our HMI colleagues, if not perform joint research.

It is safe to consider the multimedia field as mature, and as such, opportunities for consolidation exist and should be identified in the workshop for further study. Having said that, I believe opportunities also exist for expansion. A recurring theme in internal discussions where we asses our roadmaps and strategy - again our approached is heavily influenced by the need to achieve successful technology transfers into products - is the need to take a holistic view of the problems and opportunities described above. In particular, we have initiated an activity designed to look at multimedia content as an entity with a lifecycle associated with it, and a series of business process that dictate the evolution of this cycle. We believe that understanding the way multimedia content is created, manipulated, and ultimately discarded should help identify and motivate new views for existing areas of research, if not become a research topic of its own.

Thomas Plageman (U of Oslo)

The times when the keyword “multimedia” has been hype and indicated a hot research area are over. From my point of view, this is probably the biggest success of multimedia research, because the reason that multimedia is not special anymore is that everybody is using it (to a certain extent). In other words, multimedia has become for many people already a natural part of their lives. However, this does not mean that all problems are solved. Obviously, there is and will be a strong request for better quality as well as the request for using multimedia everywhere, e.g., on the road on cell phones and PDAs as well as at home in home cinema systems.

1. Your view of the multimedia field - what are the fundamental topics/areas?

Multimedia itself is not an application. Therefore, the seamless integration of multimedia in applications and the corresponding system support are important for higher quality (multimedia) applications. At the application level we have still many open issues on how to use in an optimal way multimedia to improve the efficiency and productivity of the user, how to design the appropriate human-computer interfaces and how to evaluate the perceived quality of an application. The last aspect leads to the more general question of how we should benchmark multimedia applications and systems. Can we develop standardized benchmarks that allow us to better compare related work?

At the systems level, data of many different types have to be handled. Earlier research on multimedia systems elaborated the fact that multimedia applications have been mainly concerned with retrieval and transmission of continuous media, e.g., streaming video. This type of application allowed to predict the future behavior and to design special systems solutions that utilized the knowledge about the future behavior. However, multimedia is and will not be only streaming media. Complex presentations encoded for example in MPEG4 or SMIL documents provide many interaction possibilities and reduce the predictability to a certain part. Furthermore, we will also see the trend that we are not only using read-only multimedia applications respectively write once and read many times. All these developments show that it will be in the future not anymore the best solution to design a specialized system for supporting multimedia applications. Multimedia data have to be handled in systems together with non-multimedia data (why should we make in the future this distinction?). The integration at the system level is still an open and important topic.

The question of how to design integrated systems imposes another more fundamental question – which is not a pure multimedia research question: which type of abstractions should we use to design and build multimedia systems? The layer concept has been applied successfully for many years, but it has been shown already a long time ago that the layering principle is not well suited. Other abstraction mechanisms that are studied are object-oriented technology, component technology, aspect-oriented programming etc. Do these technologies provide the needed properties for development of multimedia systems or do we need to develop better technologies?

2. What multimedia problems have been solved? What successes can research in this field claim?

The most important impact of successful multimedia research are probably useful tools and standards, like video streaming and players, the MPEG family, SMIL etc. It should be noted that most of these tools have been provided as public domain on the Internet and have stimulated both (1) a broad usage which stimulated commercial player to follow this success up, and (2) research on improvements of these tools.

As I said before, the biggest success of multimedia is that it has become already for many people a normal part of live.

3. On what problems should researchers be focussed?

There are still so many open problems and I cannot identify one problem domain that requires clearly much more attention to advance the entire research in multimedia.

4. What is the state of SIGMM?

For me, SIGMM is mainly visible at the ACM Multimedia conference and by it's web page.

5. What is the state of the annual ACM Multimedia Conference?

As a PC Co-Chair for this year's conference, I can state that we have been quite happy with the submissions to the conferences such that we could perform a strong selection process to present only good papers at the conference. For the application track, I have to say that we could improve in attracting and integrating more researchers from other related areas, like HCI, VR, etc.

Both, the workshops and the doctoral symposium are important and successful parts of the conference.

6. What changes might SIGMM pursue?

There had been discussions about multimedia curriculum and it would be good if this could be followed up in SIGMM such that all SIGMM members finally could benefit from such curriculum suggestions. To go one step further, the SIGMM might even try to stimulate the exchange of (multimedia) course material for multimedia courses. I think the latter has also been discussed before.

Lawrence A. Rowe (U of California, Berkeley)

1. Your view of the multimedia field - what are the fundamental topics/areas?

Multimedia encompasses any application that uses multiple media, but it has to be more than text and images. For me, the defining characteristic is dynamic behavior. This behavior can occur either in the data (e.g., continuous data such as audio, video or animation) or in the user interface (e.g., invoking dynamic action by activating a control interface).

Fundamental topics range from hardware architectures for capturing, processing, transmitting, and displaying or playing data to software architecture for processing, searching, or presenting multimedia data. Data abstractions to represent multimedia data and languages to specify synchronization, presentation, and user interaction behaviors are also important topics. Lastly, multimedia applications can be partitioned into various categories including: i) playback applications, ii) synchronous collaboration and tele-immersion applications, iii) content authoring and management, and iv) content retrieval. An important area in some of these applications is the use of time (e.g., latency, synchronization, etc.). Distributed and collaborative applications are much more challenging and difficult than applications that involve just one user and computer (e.g., DVD playback on a PC).

2. What multimedia problems have been solved? What are the successes?

I believe the trade-off's and design alternatives for hardware and software architectures to support media capture, coding/decoding, transmission, and display are well-understood. Some interesting challenges remain in the way data is presented (discussed below) and the possible use of multiple-processors per chip hardware architectures, but CPU architecture with DSP and pipelined special-purpose instructions (e.g., Intel MMX) are widely available.

Synchronization, both course grain (e.g., image displayed in response to user input or action trigger by time) and fine grain (e.g., audio/video sync, turn-taking in collaboration, etc.), are understood.

Real-time software and network protocols are understood, although not widely deployed. Audio/video coding algorithms are excellent. While it is likely that effort will continue on new algorithms and standards, the improvement over existing algorithms is likely to be small relative to the effort required to realize the improvement. It is quite amazing how good streaming video and audio are at 250-1,000 and 5-128 Kbs, respectively.

3. On what problems should researchers be focused?

My bias is to work on applications that solve real-world problems or that encourage wide-spread use by all users. While there are many examples of such applications, here are a few problems I think are important:

- Authoring – it is much too difficult to create non-trivial multimedia titles. Examples include a multimedia history and discussion about an event (e.g., the development of wireless communication and broadcasting technology), interactive teaching material (e.g., a simulation of a physical system), or an on-

line title created during an event (e.g., capture/display of statistics during a sporting event with the ability to go back and review selected events during this event or related events).

- Collaboration – humans want to communicate and participate in shared experiences. Telephone conversations and live performances (e.g., sports, music, plays, movies, etc.) are ubiquitous. However, telephone conversations are audio-only and primarily one-to-one, and live performances are typically passive experiences whether viewed in person or remotely on television. Video conferencing and tele-immersive experiences present many challenges including technical and social problems.

  Personally, I think controlling the production of a live event or collaboration is a difficult and important problem that, if solved, will reduce the cost of production and improve the quality of the experience. The goal should be to produce customizable experiences with television quality at minimal cost.

  The Access Grid (AG) is an emerging collaboration standard for n-way, multiple stream interactions. An AG node can be for one person (e.g., an office) or a group of people (e.g., a small or large room). A room-level node uses multiple projectors to provide a large continuous space for viewing remote participants and shared applications. Today this shared projected space is implemented by a multiple-headed graphics card in a Windows/Intel PC. This architecture causes a problem because all display information must pass through this single PC, and more importantly over one PCI bus. Some projector manufacturers are beginning to put their projectors onto the network. A significant improvement can be achieved by using a high-speed LAN (e.g., Infinibus) to interconnect PC's with various services to the display subsystem. But this will require significant changes in the graphics processing architecture and UI interactions.

- Edutainment – reading a magazine, newspaper, or book; watching television; listening to music; or playing a single-user video game consume a major percentage of time for people in modern industrialized nations. Surfing the Internet and playing computer games has overtaken television viewing for Americans under 25 years old. I believe the SIGMM research community has had little or no impacted these activities. Why not?

- Content search – many people want to find information from the great morass of data in the world. Multimedia retrieval has made essentially no impact on this problem. Remember the web is primary text and images, and the vast majority of web searches are based on text query because more sophisticated questions involving multimedia data cannot be answered.

4. What is the state of SIGMM and the ACM Multimedia Conference?

I am biased since I am Past Chair of SIGMM and General Co-Chair of MM'03. Having said that, I am worried that the research published at the annual conference rarely has an impact on real-world users or applications. All too often papers present a mature result (i.e., an algorithm that improves on current practice by a small amount or that solves a problem that no one will deploy or use) that is easy to evaluate rather than papers that

present significant breakthroughs, innovative (crazy?) approaches to solve challenging problems, or raise questions that will have real impact.

We should seriously consider breaking the current conference up into a series of mini-conferences that run in parallel on a broader array of topics. We should also consider re-structuring the conference to increase the number of posters so that more results can be presented and discussed. But, both actions must maintain the important quality metric for published papers. Perhaps a second conference should be organized each year that is dedicated to a particular subset of topics presented only as posters.

Some other problems with the current conference are: i) few people other than presenters attend the conference, ii) practitioners (e.g., content authors) are poorly served by the conference, iii) tutorials are poorly attended, and iv) a multimedia show and exhibition should be held in conjunction with the conference. But, as with the SIGMM changes discussed below, the problem is getting volunteers to do the work.

5. What changes might SIGMM pursue?

SIGMM should do many things including: i) resolve the Multimedia Systems Journal issue (i.e., make the journal a first-class ACM journal included in the digital library), ii) develop a multimedia curriculum for colleges and universities, iii) encourage the development of new content for the ACM Digital Library, and iv) provide more benefits to the community. The goals should be to maintain quality, increase services, and enlarge the community.

The primary limitation is volunteers to take responsibility for making changes.

Henning Schulzrinne (Columbia University)

1. Your view of the multimedia field - what are the fundamental topics/areas?

In this note, I will focus on continuous, distributed multimedia services, including media-on-demand (streaming audio and video) and interactive media such as video conferencing and Internet telephony.

Networked multimedia dates back to at least the early 1990's, if we take experiments involving audio and video crossing wide-area networks as the relevant marker. (The earliest experiments in voice transmission over the Internet date back to the 1970s, but there appears to have been little research progress between those early experiments and the re-awakening of interest in the late 1980s.)

Networked multimedia clearly has established itself as a major Internet application. (Caveats on this success are noted below.) By any measure, this indicates that the research and development efforts in this area have paid off.

While taking at least a decade later to make significant dents into the existing circuit-switched phone system, the transition is no longer a question of "if", but "when". It appears that research, development and deployment on the circuit-switched infrastructure has largely come to a halt, with the possible exception of the filling-in of some second-generation wireless networks. Major U.S. long-distance networks are converting to packet-switched technology, at least in the core, and initial consumer deployments and business roll-outs of voiceover-IP (VoIP) are picking up speed, albeit with increases measured from a very low baseline. On some price sensitive and high-volume international routes, VoIP already commands a significant fraction of the traffic, even if hidden behind cheap prepaid phone cards. At this point, deployment appears to be primarily paced by the rollout of residential broadband services and the very long depreciation intervals for PBXs and carrier switching equipment.

Networked multimedia had to solve the following technical problems to reach that state:

- high-compression, high-fidelity codecs;
- operating system interfaces;
- packet loss compensation;
- playout delay compensation;
- multimedia transport, including security mechanisms;
- signaling (session establishment);
- interworking with legacy systems, primarily for VoIP.

While there is probably progress to be made on all of these topics, it appears that many have reached the point of diminishing returns, coming within a close margin of the optimal solution or where significant improvements would require disproportionate amounts of computational effort or complexity.

One could argue that IP multicast should be considered an enabling technology, but its use is so uncommon for multimedia applications, that it is hard to consider crucial.

2. What are the failures?

Failure: prediction and estimation errors

We can identify two core failures related to the research and development of networked multimedia applications: overestimating the speed of deployment and the importance of interactive video. During the early days of Internet multimedia, the consensus was that interactive multimedia would completely dwarf existing traffic types – at the time, file transfer and email. This clearly never happened, with continuous media probably consuming less than 5% of Internet backbone bandwidth. There are at least three reasons for that: First, most video is watched at home, and residential broadband has barely become available. Secondly, for real-time viewing of high-profile events such as sports, the existing TV is hard to beat in terms of quality and viewing experience. Thirdly, for recorded entertainment content, shipping DVDs by mail and file transfer within peer-to-peer networks offer far larger selection and, for Netflix-like services, lower bandwidth cost than audio or video on demand.

Like most researchers, the network research community focused on progress and promise of progress in its own field, ignoring the continuous dramatic reduction in per-byte cost of disk storage. This made storing content on portable devices and PCs far more attractive than streaming it in real time.

Researchers probably also ignored the fundamental difference between, say, the web and streaming media applications, leading them to grossly overestimate the displacement speed. While the web offered a fundamentally new way to obtain information, continuous media applications typically tried to replace existing dedicated networks and systems such as TV and radio, and usually rather poorly in terms of content quality, end system cost, reliability and ubiquity.

For Internet telephony, the significant decrease in long-distance rates made Internet telephony far less attractive and limited its initial appeal to the segment of the population with both technical expertise, a tolerance for low quality and reliability, a lack of financial resources and spare time on their hands. This made international graduate students a prime consumer of such services, but does not translate readily into a mass market.

The network research community grossly underestimated the difficulty of deploying new network-layer services such as IPv6, multicast and resource reservation. All of these have now been available for around a decade and have seen no significant deployment. (IPv6 appears likely to see such deployment over the next decade, but only if 3G networks are not stillborn.) The community focused on the purely technical aspect of scaling, declaring point-to-multipoint conferences as primitive and wasteful, but not recognizing that the deployment and management difficulties of IP multicast far outweighed the largely user-invisible costs of inefficient packet replication.

Repeating the mistake made since the first demonstration of video phones, researchers, despite warnings to the contrary, continued to overestimate the value of video in conferencing and often assumed that almost all interpersonal communications would be

conducted by video. On the other hand, the notion of text-based messaging received almost no attention, even though it has become far more important than interactive video communications.

Failure: research tools

Despite about ten years of active research, there is still a lack of common, state-of-the-art interactive multimedia tools that are suitable for experimentation, teaching and as a development platform. Tools like *rat* and *vic*, while initially providing such a platform, have stagnated in the past few years, with little active maintenance and an architecture that is not particularly amenable to adding new features. Other tools and libraries, such as the Java Media Framework, continue to be brittle if used outside the dominant Windows platform and suffer from excessive delay and other limitations. This lack of tools makes it difficult to deploy large-scale systems, such as within the context of Internet2, and thus move beyond the classical demo environments.

Failure: current practical challenges

Writing interactive multimedia applications that are robust in the face of network and OS impairments and offer low latency continues to be more difficult than necessary.

Even on the dominant OS platform, low-latency audio and video APIs are difficult to find, with DirectAudio adding too much delay and DirectKS being barely supported and too low-level. (The only alternative is to use a packaged system such as the RTC library.)

Little progress has been made on non-audio and video multimedia interaction, with T.120 still dominating the field, despite numerous shortcomings and the awkward fit into modern architectures.

3. On what problems should researchers be focused?

Given the continued absence of viable resource allocation mechanisms in the Internet, dealing with network impairments by adapting application behavior and adding FEC continues to be of interest, with a greater emphasis probably on media-type optimized behavior rather than generic algorithms.

In general, as the community realized that network services for multicast, mobility and IPv6 in general may never be deployed universally, there has been an interesting shift to see how application-layer techniques can be used to achieve similar results, with more incremental deployability. This motivates research in application layer mobility, routing and multicast, and peer-to-peer research for continuous media.

For many years, the central mantra of network design and applications was probably “anywhere, anytime, any media”. Now, the real-life reaction to the fulfillment of this goal has been to request that people turn off their cell phone. Our communication control mechanism are still fairly dumb, with research often assuming that communication is always desirable. While the notion of ubiquitous computing has also traditionally emphasized universal access, I believe that a realization has set in that controlling reachability and managing access has become at least as important. This topic of context-aware multimedia communications does not seem well represented within SIGMM events.

Down-scaling traditional concepts of multimedia communications to devices that have roughly the computing power of a 1992 workstation, but are also bandwidth and power-challenged, continues to offer challenges, particularly if combined with opportunistic networking, i.e., the ability to use whatever network services happen to be available, whether 2G, 3G or wireless LANs.

4. State of SIGMM and the ACM Multimedia Conference?

Based on my personal experience, I will comment only on ACM NOSSDAV (Network and Operating System Support for Digital Audio and Video), a conference that has been supported by SIGMM for the past few years. While it continues to attract around 50 to 60 papers a year, attendance at the conference has been decreasing. Similarly, IEEE Infocom 2003, for example, only offered a single multimedia-related session.

It may be helpful to investigate joint events with the pervasive, mobile and wearable computing communities, as many of the challenges in multimedia communications seem to have found a home there.

Ralf Steinmetz (Darmstadt Univ. of Technology)

1. *Your view of the multimedia field - what are the fundamental topics/areas?*
   - The fundamental areas of the multimedia field are still the traditional areas of topics that are related to audio and video in combination with other media. Many "sub-areas" like video encoding can be derived (to some extent) from this area.
   - One significant characteristic of multimedia data is the fact that it has a high resource consumption. Larger space is needed for storage; higher bandwidth is needed for transport.
   - In relation to data transport timing constraints are an important topic.
   - One major problem for the SIGMM is the fact that many of the MM related topics are also topics in other SIGs or conferences. VR for example is a major topic at SIGGRAPH. Transport related issues for MM data or application are also a topic of SIGCOMM.
2. *What multimedia problems have been solved? What are the successes?*
   - MM has become successful in the Web. Many pages in the Web offer multimedia content.
   - Multimedia as part of learning applications in the sense of TELElearning (and videoconferencing) can be seen somehow there as a solved problem and has also lead to the success of these applications
   - Partly solved is video streaming because it is a problem that is related to more than one medium. Content analysis, for example, is usually dealing with only one medium, either video or audio, but not with the combination of both. Thus, there are many problems, which can be regarded to be solved in relation to a single media.
3. *On what problems should research be focused?*
   - Research today is still mainly focused on isolated topics like coding.
   - Popular trends in communication systems and in the research community that is concerned with such systems are seamlessness, ad-hoc/wireless or peer-to-peer infrastructures, mobility and ubiquity. Research should focus on problems that regard the aforementioned topics in combination with multimedia.
   - Management of MM data. Recently the work presented at the ACM MM conferences was not concerned with content management. The amount of multimedia content will increase rapidly in the near future which will make the management of this content necessary. One could rather say that the focus should be on "multimedia content management systems" since there is a large amount of existing work on CMS systems for web documents.
   - Applications that were not in the focus of the MM conference were games and applications for the production of MM content. Games are a very interesting

application because multimedia content like audio, video and VR environments must be orchestrated. In the case of distributed games also problems related to the transport of MM data become interesting. Only little attention was given to applications in the e-Health area.

   - The production of MM content is mainly performed manually. Not many automated systems are available and almost no work has been presented in recent conferences or issues of the MM Systems Journal.
   - Metadata in combination with multimedia should be a future research topic. Management of MM data can be performed much more efficient with the aid of metadata. Almost all MM applications could benefit from the information that is provided by MM metadata. Research topics in this area are, e.g., automatic creation of metadata or the enrichment of metadata with the aid of ontologies.
   - What about synchronization of MM data? Many problems are solved higher capacity links and power of the clients (e.g. sufficient RAM to buffer data). New problems might arise with small, mobile devices with scarce resources.
4. *What is the state of SIGMM and the ACM Multimedia Conference?*
   - ACM Conference: OK to me in terms of quality. Could be larger attendance
   - SIG MM: operates just for the conference, if more is wanted, who shall do the job?
5. *What changes might SIGMM pursue?*
   - Topics like seamlessness or ad-hoc infrastructures were not mentioned in the CFP for the ACM MM 2003 conference. A discussion about future trends in multimedia should lead to innovative topic, which should be included in next years CFP.
   - The areas of interest mentioned for the MM Systems Journal should also be discussed and refined.

Michael Vernick (Avaya)

1. What are the fundamental topics/areas in the field of multimedia?

Computer software, systems and networks for the creation, management, transport and presentation of digital video, audio and images.

2. What multimedia problems have been solved?

I don't know if there are any that are completely solved, but the problems which are mostly solved might include: Video, audio and image compression, large-scale video servers, video shot indexing, single media authoring tools.

3. What successes can research in this field claim?

In general, we have been successful in maintaining the distinct field of Multimedia.

We have been successful in defining multimedia standards such as RTP/RTSP, MPEG and SMIL.

Also, Audio streaming on the Internet and voice over IP have matured into stable, decent quality products.

4. On what problems should researchers be focused?

- Understanding the needs and problems of product oriented companies (for shorter term/applied research)
- Systems that incorporate research from several areas
- Multimedia digital rights management and security
- User interfaces and new metaphors for multimedia communication
- Multimedia for heterogeneous end-user devices
- Multimedia conferencing on the Internet with audio and video
- Monitoring the transmission of video and audio across networks. For example, if I pay $X to see a live sports event but the quality is poor, the reason for quality degradation should be known to both the producer and consumer. In a perfect world, the quality of data flowing over an IP network would be the same as the plain old telephone network.

5. What is the state of SIGMM?

I don't currently see the advantages of joining the SIGMM, other than receiving the conference proceedings. Membership does support the community and the overhead for the conference but I think that we could have more interaction possibilities for members.

What is the state of the annual ACM Multimedia Conference?

Stable. Not shrinking, but not growing. If we are happy with the current state, then no major changes are needed. On the other hand, do we want to be more like a SIGCHI or SIGGRAPH? If we are interested in becoming a bigger conference and organization I believe that we need more industrial participation, not from just industrial labs, but from

influential companies like Apple, Real, Microsoft, Macromedia, etc. Also, most of the people who attend the conference are involved in some way with the conference, (presenters and committee members). We don't seem to get many people not directly involved. How do we attract outsiders?

What changes might SIGMM pursue?

- More compelling reasons for why someone should join the SIG.
- A web site that lists descriptions of projects that members are working on.
- A monthly online seminar using WebEx or something similar where members present current research projects
- More industrial participation in the conference with possible separate review criteria

Harrick Vin (U of Texas)

Multimedia Systems Research: A Retrospective
OR
Whatever Happened To All That QoS Research?

Over the past decade, techniques for providing Quality of Service (QoS) guarantees to applications have been the central theme for much of the research in multimedia systems. This broad category of research includes techniques for designing multimedia servers (VoD servers), network protocols and scheduling techniques for per-hop and end-to-end QoS guarantees, and QoS mechanisms for operating systems, among others. Several thousand papers have been published in these areas. Yet, few of these techniques are deployed in real systems today. Even though we understand a lot about designing systems that can provide QoS guarantees, few systems today provide any form of service differentiation or guarantees.

So, the research community should get an "A" grade for trying hard; however, I contend that the community should get a "C" or a "D" (at best) for the impact it has had on practice.

Hence, the natural questions are: what happened to all that research? Why has it had such little influence on practice? What can we learn from our experiences? It is my belief that our community should perform some introspection to answer these questions; understanding these issues will have significant impact on defining research agenda for the next decade. In what follows, I attempt to identify some of the reasons for our lack of success.

The lack of impact on practice, in my opinion, is a result of many factors. The lack of a "business case" (or charging models) is often cited as a reason for the lack of QoS support in systems today. Although this certainly is one of the reasons, it, by no means, is the only one. The following are a few other reasons to consider.

- Much of the QoS research has become victim to the "Moore's Law". Much of QoS research attempts to manage resources carefully with the intent of providing service guarantees to applications. This assumes that the resources available in the system are scarce, and must be used judiciously. This assumption is often true at the time the research is conducted; however, the community has rarely asked the question: how technology dependent is the problem (i.e., will this remain a problem in a year or two)? What level of over-provisioning will solve the problem (and how much will it cost to over-provision the system)? What is the complexity of careful resource management and how long will it take to deploy such solutions? Will the application demands keep pace with technology improvements? Although there are special-purpose applications where careful resource management is essential, for most applications, we have seen that relatively small amounts of over-provisioning eliminate the need for careful resource management. Hence, techniques for providing QoS guarantees have often lost to Moore's law!
- As a community, we have done a poor job of demonstrating the significance of the problems prior to solving them. We have often lacked applications as well as workloads for justifying or evaluating our research. We have little understanding of what QoS guarantees do applications/users really want; we have rarely argued/demonstrated that the operating regimes where QoS support is beneficial do in fact occur (or likely to occur) in practice. For instance, much of the research often assumes an operating point (e.g., high-level—90% and higher—of resource utilization); however, most service providers (network, computation or storage) rarely run their resources at these levels of utilization!
- We have paid little attention to deployment considerations for our solutions. Much of the QoS research assumes an "all-or-nothing" model for deployment—we have rarely asked the question: can these techniques be deployed and utilized incrementally? What are the benefits of partial deployment? Are the solutions backward compatible (how will applications developed without the knowledge of our wonderful solutions work in the new environment)? Lack of significant benefits with only partial deployments and lack of backward compatibility make it difficult for service providers to begin deployments—hence, the all-or-nothing model often fails!
- We have often solved intellectually challenging problems in designing systems that can provide QoS guarantees. However, it is often the case that actual deployments of these solutions require several "magic" numbers or help from other components (e.g., applications specifying QoS requirements). We have rarely asked the question: how difficult is it for applications/users/system administrators to deploy and use these solutions?

So, what can we learn from all this? Here are a few things to consider.

- I contend that Moore's law will help solve many of the performance-related (e.g., delay, bandwidth, loss, etc.) QoS problems for the most common applications. Careful resource management will be required only in special cases.
- Most of the systems (multimedia and others) today are too difficult to configure, manage, and use. The research community should perhaps focus on "QoS" areas such as availability, security, manageability, deployability, etc. – rather than performance issues. Making progress in these areas, however, is likely to be challenging. Today, we do not even understand how to formulate there problems precisely or evaluate solutions that address these issues!
- As a community, we should try to define benchmark applications and workloads that can provide guidelines for our explorations and evaluations. We need models for how applications, workloads, and technology will evolve—these models together will help identify fundamental problems.

In summary, it is time to evaluate our progress and develop an agenda for the future. Let's make sure that we learn from our successes and failures. Let us make "impact" as a primary criterion for developing and evaluating a research agenda. Let us develop simple, incrementally deployable solutions. Also, let us try to develop techniques that can transcend technological and applications boundaries.

This will be hard – but then who said that getting an "A" for impact will be easy?

Lynn Wilcox (FX PAL)

From certain points of view, multimedia is a mature research area. We can watch a DVD and skip to a particular scene or follow a hyperlink to obtain related information. We can view streaming video on the Web from either live events or from pre-recorded data. Live seminars and classes can be broadcast along with the presentation material. These seminars and classes can also be recorded, and made available on the Web for later viewing. We can upload video and images from our digital cameras and create multimedia shows with music and titles. However, based on the use experience of FXPAL multimedia technology, we found that multimedia is not a solved problem. Although our current technologies are useful, they do not always solve the problems we think they do and do not always provide a complete solution. Areas we found that need more research include note-taking, hypermedia, adaptive media analysis, automated capture, annotation, interfaces for small devices, reuse, and authoring.

One experience we had with multimedia use involved the Stanford University Medical School. We used FXPAL technologies to support indexing and summarization of a pathology class for Winter Quarter. The University was already recording lectures and putting them online for later access. We thought that the indexing and summarization technology we had developed at FXPAL would help students to better learn the material. Our hypothesis was that our indexes and summaries would help students find specific information in the lectures without having to watch the entire video, and that this would make their learning more efficient. What we found by holding focus groups with students and professors and examining video logs was quite different. Students mostly used video to watch a lecture they had missed, and typically watched it at double speed. They used our indexes, but mostly as a "smart" fast forward. They liked the summaries, since they provided a quick overview of the lecture as a memory aid. However, they did not make full use of our technology for information mining, as we had anticipated, and did not think improvements to indexing and summarization would help them. On the other hand, during the focus groups they made a number of comments that suggested alternate research paths. The students wanted the ability to create links between the various sources of information provided to them: notes, recorded lectures, handouts, class syllabus, and textbook. They felt, as did their professor, that they learned by correlating diverse sources of information. Thus note-taking, annotation, and hypermedia may be promising research areas for this type of application.

Another use experience came from a product for multimedia management that our parent company sells in Japan. The product is based on the indexing and summarization technologies developed at FXPAL. As they began to market this product to diverse customers, they began to discover problems. One was the ability to handle different genres of video. Our indexing and summarization algorithms were based on shot changes, but some videos had only a single shot, while for others our algorithm detected far too many shots to make the indexes useful. Although in theory our analysis algorithms could be trained to work on different types of video, it proved too complex to gather the data required for training. Another problem they discovered was that many of the customers did not have the ability to capture their multimedia content. And finally, some customers needed to access multimedia from PDAs. This feedback suggests research in genre-

adaptive analysis techniques, automated media capture systems, and interfaces for small devices.

In order to gain an understanding of the use of multimedia archives, we performed an experiment with Television Archive. Television Archive is an online collection of video broadcasts from around the world just before and after September 11. We applied FXPAL indexing and summarization technologies to this collection, and demonstrated how they could be used to access specific information (in particular, video clips of the plane hitting the building). The idea of public online video collections is being extended in OneWorld TV, which allows people to upload their own video content concerning social issues and provides global access to this data. The expectation for both of these online collections is that people will want to find and reuse parts of videos for their own multimedia authoring purposes. We are monitoring use of these archives, and if these expectations prove to be correct, research in video search and editing will be needed.

Finally, many products are beginning to emerge that allow users to manage and share their personal photo collections. There are also tools to help people edit home videos. Although we are doing research in this area, we have not seen much feedback from users of these applications. However, it is certain that they will want technology that is both useful and usable.

Multimedia technology is being used today in a number of applications. By studying usage, we can learn productive directions for multimedia research and thus improve existing technologies and build new and more useful ones. Our use experience suggests research in note-taking, hypermedia, adaptive media analysis, automated capture, annotation, interfaces for small devices, reuse, and authoring. The HCI community has expertise in studying how people use technology. It might be profitable to extend the scope of the ACM MM conference to include multimedia use and practice.