Panel

New Challenges in Multimedia Research for the Increasingly Connected and Fast Growing Digital Society

Jia Li[1] (Chair), Shih-Fu Chang[2], Michael Lesk[3], Rainer Lienhart[4], Jiebo Luo[5], and Arnold W. M. Smeulders[6]

[1] The Pennsylvania State University, USA
[2] Columbia University, USA
[3] Rutgers University, USA
[4] University of Augsburg, Germany
[5] Eastman Kodak Research, USA
[6] University of Amsterdam, The Netherlands

ABSTRACT

We are living in an era with rapidly growing connectivity through digital networks. Multimedia research will continue to address crucial problems in the digital world, such as image/video/audio retrieval and tagging. On the other hand, new challenges and opportunities arise when we accumulate multimedia data and expand Internet social networks at astounding speed. This panel, formed by active researchers from both academia and the industry, discusses potential ways to enhance the management, sharing, and archiving of multimedia objects in socially networked environments, and reversely, ways to reach computational goals made possibly more achievable by the abundance of data and user connections. This paper records the opinions of the panelists.

Categories and Subject Descriptors

H.3.3 [**Information Storage and Retrieval**]: Information Search and Retrieval

General Terms

Algorithms, Design, Experimentation, Human Factors, Performance, Standardization, Theory

Keywords

Image Retrieval, Video Retrieval, Social Networks, Computational Photography, Computational Art

MIR'07, September 28–29, 2007, Augsburg, Bavaria, Germany.
ACM 978-1-59593-778-0/07/0009.

1. INTRODUCTION

Digital multimedia collections are growing at an extraordinary rate. With increasingly affordable digital data acquisition equipment (e.g., cameras, camcorders, cell phones, voice recorders) and the widespread use of personal computers with abundant storage, users are generating digital objects on a daily basis.

Large-scale online multimedia sharing communities are being created or expanding. According to recent reports, an Internet photo-sharing startup, flickr.com, has several million registered users and hosts several hundred million photos. New photos are being uploaded in the order of one million per day. Founded in early 2005, YouTube has already emerged as a global Website traffic leader, delivering more than 100 million video views and adding 65,000 new videos daily. Other fast growing social network and multimedia sharing Websites include MSN Live, MySpace, Orkut, Wikipedia, and The Facebook. More specialized online photo-sharing communities, such as photo.net, deviantART, and airliners.net, also have images in the order of millions entirely contributed by the users.

Comparing with conventional pathways of publication or personal communication, the Internet is unique in multiple ways. Modern software has made it easier to create multimedia documents with impressive appearances. On the other hand, multimedia pieces visited by even millions of people on the Internet can be of rather informal format. The procedure to expose one's creation through the Internet, let it be an article, music, video, or art work, is simple and nearly cost free. The creator also has the convenience of not revealing his or her true identity, and hence, with less concern about being judged, can express more willingly and freely. These special traits of the Internet open unprecedented opportunities for individuals to showcase their creative minds, to harness the expertise of numerous online users, and to socialize on large scales and with enormous efficiency. The magnetic effect of the Internet is evident from the quickly expanding blogs, video/picture sharing sites, and social network sites.

Multimedia researchers will continue to face the challenge of intelligent management of ever increasing multimedia data. We would also like to be more proactive at inventing technologies for the public's social or educational interests. Moving beyond developing functions for the Internet, we may more often find ourselves leveraging the infrastructure and social structures established on the Internet.

This panel aims at identifying new challenges in multimedia research in response to the fast growing digital world. The opinions of the panelists are documented in Section 2. When the written comments were collected, no length limitation was set. The panel moderator Jia Li maintains a neutral stance during the panel discussion. However, in the interest of keeping a record, her personal opinions are incorporated into this article.

2. MAIN ISSUES DISCUSSED

In this section, we list the main issues discussed during the panel and the written comments of the panelists.

2.1 Personal Multimedia Retrieval

With the wide spread of very affordable digital cameras, individuals are generating enormous amount of digital photos. What research problems can be explored to help people better manage, share, and archive their personal pictures?

Shih-Fu Chang: This depends on the anticipated type of usage. For management of personal photos, studies have shown simple organization based on time and event (e.g., "trip to Beijing") will suffice in most cases. Additional useful tools might include automatic techniques for grouping photos into subfolders sharing personal memorable experiences, e.g. climbing the Great Wall, dining with old friends, etc.

One of the interesting research topics will be gaining further understanding of the dimensions in which users organize their photos into subfolders. Are there common dimensions shared among users? Or does each user have his or her own personal preferences that may vary in different contexts and change over time? Understanding of such user behaviors will be valuable for multimedia researchers. If there are common dimensions, then optimizing automatic methods for such dimensions will be critical; if the preferences are highly individual, then the focus should be on development of intuitive interactive systems that are easy to use and flexible for accommodating various user needs.

Intelligent methods for recognizing humans and human activities continue to be critical for this domain, despite the prevalent technical difficulty caused by well known problems such as variations in imaging conditions, subject appearances, and aging etc. Any tangible progress in this area will have huge impact. Imagine the functionalities it may enable - grouping photos based on persons, linking photos to personal web sites or social networks.

Automatic tagging by content analysis and classifiers is another direction that may pay off. Some statistical models have shown reasonable performance in detecting semantic concepts in images and videos, such as location (indoor/outdoor, beach, mountain etc). Such detected labels will be useful in making the photos more searchable, especially when the photos are shared with other people who are not familiar with the original settings/events of the content. One interesting research topic related to this is to discover additional semantic classes that are useful in practice and detectable with reasonable accuracy.

Michael Lesk: Let me skip over the obvious—image analysis for recognition of specific objects, GPS in cameras, and face recognition. I will also skip the general questions of digital preservation and archiving. Instead, let me suggest a few perhaps less usual directions:

1. Detection of emotion in faces.
2. High level scene characterization: How many people are in this frame, and what does that tell us? Can we recognize such general topics as "parties", "sports", and the like.
3. Social grouping: Does your picture look like anything on flickr and how are the best matches there characterized?

Rainer Lienhart: The core research problem will be how to exploit the mass of photos and its sparse meta-data (in highly diverse domains such as text, tag space, image, storage location, emails recipients, etc.) to propagate labels across most of the photos, for instance, how to propagate location information (such as GPS coordinates) to unlabeled images which obviously must have been taken at the same place.

Jiebo Luo: I think at the core of the problems is still semantic understanding of multimedia content. We do not need much argument to realize that with (rich) semantic representation and annotation, it will become easy to manage, share, archive, search, retrieve, and even enhance/edit personal digital multimedia asset, including photos and videos.

We should broaden the scope beyond digital photos to include digital video clips. With the convergence of both photo and video capturing capabilities in a single device such as digital cameras and camera phones, consumers' photo collection is becoming heterogeneous. I would argue that short video clips are sometimes more valuable to consumers for several reasons: (1) they capture more scene dynamics, be it motion related or space related (zoom and pan), and (2) each short clip is a clearly delineated event (compared to videos from a camcorder). However, video also poses a greater challenge for management. To the first order, it is not easy to browse video files in a way similar to browse photos through thumbnails. And the sheer storage requirement for video data is a big challenge in and of itself. While there is richer information and the possibility of multimodal fusion (discussed by last year's panel), I personally do not think video annotation and search are necessarily easier. In fact, with amateur videos and their unconstrained content, consumer video understanding has not been studied as extensively as consumer photos. Outside news and sports video, semantic video understanding may even be a bigger challenge.

Everyone knows we have a "semantic gap" here. We all know low-level features just do not cut it anymore, but it is not easy to get to the semantics level. Within constrained domains such as news and broadcast sports media, considerable progress has been made with coordinated efforts such as TRECVID. However, personal multimedia data are far more challenging to handle. Kodak has collections of extensive photo and video datasets designed to span the "photo space" and "video space". When we

asked some of the leading teams in TRECVID to apply their TRECVID-driven technology to our consumer digital assets, the results were not pretty. Even after some level of parameter tuning using consumer data, the performance is still no where near what they achieved with news videos in TRECVID.

It will take more than just the multimedia research community to make inroads into the consumer multimedia space. It is happening now. Many multimedia researchers are reaching out to machine learning and computer vision techniques, and many machine learning and computer vision researchers are working on multimedia recognition problems.

I would like to emphasize one particular area of research—the integration of content and context. By content, I mean the information contained in the pixels and frames. By context, I mean the information surrounding the pixels and frames. Integration of content and context is crucial to human-human communication and human understanding of multimedia: without context it is difficult even for a human to recognize various objects and scenes, and we become easily confused if the audio-visual signals we perceive are mismatched. For the same reasons, integration of content and context is likely to make (semi)automatic content analysis and indexing methods more powerful for managing multimedia data. It can help narrow down the semantic and sensory gap by a portion that is difficult or even impossible to achieve using classical approaches to (semi)automatic content-based understanding and indexing.

There are many interesting research problems in this area, such as:

1. Contextual metadata extraction
2. Models [4] for temporal context, spatial context, imaging context (e.g., camera metadata), social context, and so on
3. Web context for online multimedia annotation, browsing, sharing, and reuse
4. Low-cost context tagging systems, e.g., geo-tagging, voice annotation
5. Context-aware inference algorithms
6. Context-aware multi-modal fusion systems (text, document, image, video, metadata, etc.)
7. Models for combining contextual and content information
8. Using context to browse and navigate large media collections

In fact, we are organizing a special issue in IEEE Transactions on Multimedia on this very topic, scheduled for 2008 publication. We certainly hope more and more researchers will pay attention to this area and make breakthroughs.

Arnold Smeulders: Let me start by making a side-step. Digital cameras are one aspect of the digital picture revolution. The other ones are the following:

1. The availability of very affordable digital cameras.
2. Digital communication is widespread and available at speeds permitting easy exchange of the picture data of good resolution images and in the near future even of good resolution digital video.
3. The combination of cameras on board mobiles making pictures very close not only to the photographer but also to the ones for whom the photo was taken. A photo is just seconds away (of being delivered!).
4. The disk capacity of home computers is so cheap that storing large quantities at home is no longer a problem. It is a real possibility.

It is important to realize that the last two elements of the digital picture chain develop slowest:

1. The compute power doubles every second year, slower than any of the other factors above, e.g., affordable storage, affordable sensors, and affordable network communication speeds. This is a factor limiting a further expansion of the wide-spread availability of digital management of pictures.
2. The computational method to provide access to pictures is still a hard problem. In spite of the tremendous progress of automatic access, at the same time it is also clear how hard the general problem is.

All aspects are needed to provide people with management and archiving capabilities. To share pictures, the bottle neck is communication more than anything else and there the essential mile stones are behind us. The fundamental research needed in this area is among others:

1. Automated recognition of visual concepts. A start has been made but still a very hard problem in its general scope. Humans use one quarter of their brains to solve the problem, so it is unlikely this is a simple problem. The problem will be reached by features, machine learning and annotation of real-life examples. All three aspects can still be improved. The issue is subdivided in
   (a) Machine learning methods (of active learning very many classes)
   (b) The features invariant yet powerful
   (c) Schemes for efficient active annotation of large sets of examples
2. The management of uncertain labels intrinsic to manual or automatic labeling of pictorial data. This is both of importance to pictorial database systems, the combination of social labeling as well as to the combination of evidence in automatic access.
3. Compression schemes which are built from semantic clues rather than blocks of data. Currently JPEG blocks may be disastrous for the quality of subtle features and certainly depending on where the feature is: in the middle of a block or at the edge thereof. In other words, in all current aspects of management of digital pictures, improvement is necessary.

Jia Li: With the ease of creating digital photos, I sometimes feel overwhelmed by the abundance of my own digital picture collection. I usually store my photos as one batch per trip. Without any organization, it is difficult to find pictures or recall the highlights of the trip not long afterward. Software that performs automatic or semi-automatic categorization of pictures, for instance, into landscape, wild-life, portraits, events, can be helpful for better management of personal archives.

Another potentially useful research problem is summarization of pictures, i.e., to select representative or favorite images from a given set. A related work is the story picturing engine [2], where representative images are selected based on story description. The core technique used can be useful for image collection summarization, but there is ample room for improvement. Different users may have different standards on representative pictures. It is interesting to learn user preferences and adjust photo organization tools accordingly.

2.2 Amateur Photography

Amateur photographers are increasing. Can multimedia research help people enhance their photography skills or make it easier for people to learn from experts or each other?

Shih-Fu Chang: This is happening already - in the cameras and latest photo editing tools. Motion stabilization, super-resolution, image de-blurring, and many post-processing techniques have been actively studied in recent works presented in CVPR, ACMMM, and ICIP.

Besides the above post-processing tools, research on auto albuming and multimedia slide show has shown very interesting results. Such tools greatly enhance user's capability to "tell stories".

On the other hand, I don't think users' behaviors or skills of photography should be changed. The goal of our research is to make the whole process of experience capturing and story telling much easier and as transparent as possible.

Michael Lesk: According to the Statistical Abstract of the United States, in 1992 12% of the adult population said they had "engaged in photography" in the last twelve months. In 2005 (table 1225) it was 11.8%. Of course, the population is increasing, so I guess we do have more photographers. Anyway, to answer the question I believe in a perhaps strange idea, which is that photographs could be scored automatically for quality, because many photographs for which image recognition is hard (out of focus, poor lighting, subject obscured by intervening objects) are also often going to be bad photographs aesthetically. Probably more successful would be the use of image searching to let one find other pictures of the same subject or the same style, and looking at these examples might improve one's own work.

Rainer Lienhart: I expect that in the future cameras will "fake" images to make them look better (such as Futji came up with deep blue sky despite the sky was not that blue on that day). Sharpening is already such an operation. In future, software might remove many artifacts. They might also indicate that the lighting is wrong and that the photographer will have to move to a different location. Of course, only professional will accept these kind of instruction. The "normal" people just want that the camera fixes everything by post-optimization (I will call it "faking the image").

Jiebo Luo: This is a very intriguing question. The answer is definitely. A friend of mine told me many years ago that he was a frequent visitor to Kodak.com because the Website offers rich tips and tricks on how to take the best pictures. Another story came from my own experience: I was working on the automatic cropping algorithm and discovered just how often the average consumer would frame the main subject right in the middle of the picture. My algorithm certainly took advantage of this fact but we all know a good picture takes sophisticated composition, among other things. By the way, many users liked our cropped versions better than their originals. We can help them make photography more fun and more enjoyable.

Multimedia search can facilitate the sharing of photos, and in an implicit or explicit way, improve people's photography skills. Imagine a user uploads a bunch of his pictures taken at the Grand Canyon. Through multimedia search (either by image content, GPS, or user tags), we can link the user to other people's photo collections taken at the same place, plus pictures taken by the professional photographers. The user may not know how to take a great picture, but he can recognize one when he sees it. Something very powerful can happen - he can learn even subconsciously how to take a better picture by looking at the better pictures. Of course, we can make it more educational (without being belittling) by offering suggestions in a general (what are good practices and with examples) or individually tailored fashion (what can be done better and with examples). The latter is harder to do but should be an interesting research problem. I think a Website that can provide this kind of service will draw good traffic and loyal following.

Arnold Smeulders: Essentially not. Yes, of course one could compose a multimedia learning book on how to improve photography and this would be more efficient and reach more people than common books. This is presumably not what the question is asking. Yes, of course one could also measure the technical quality of the picture and even assign some values to the aesthetics of the picture, but it would have no effect on the overall quality of photography as photos which stand out are just a few by definition. Hence my answer: essentially not (and I am not sure we should try apart from writing good books and performing some basic checks).

Jia Li: There has been some recent research devoted to computerized aesthetics ranking. This is a highly challenging research problem although limited progress has been achieved [1]. Some may argue that aesthetics is a very subjective judgment and the computer can hardly be useful in this matter. There is also the subtle issue of drawing the line between good quality and good aesthetics (or can we?). I believe it is possible to benefit from a computer's "opinion" on aesthetics, especially if the computer can pin down the shortcomings of a picture when low aesthetics is predicted. It is usually difficult for amateur photographers to obtain experts' suggestions on ways to improve a photo. In online photo sharing sites, comments given by visitors are often expression of feelings such as "beautiful", "boring", but rarely constructive. Because of the subjective nature of aesthetics, user study will be especially important for the design of any computerized system that offers aesthetics evaluation as well as the practical deployment of such systems.

When I go to a national park, one of the first things I do is to check out poster cards at a local shop. I would like

to get an idea about photography opportunities at the site. Ideally, my digital camera could have a built-in function that allows me to see highly rated pictures of well-known places. Actually, this function should not even take a lot of storage because thumbnails are sufficient. Networked mobile devices can achieve the same functionality, but they may not be connectible in remote places and it takes time to connect and navigate.

Content-based image retrieval can help photographers find similar pictures and learn from others' photography styles and ideas. Conventional content-based retrieval systems emphasize finding pictures of close semantics. This may not be the primary interest of photographers. They can be more concerned about composition ideas, for instance, in terms of color, tone, shape, or perspective. Can image retrieval systems be developed to address these aspects? In fact, certain composition characteristics can be easier to capture using low level pictorial features than semantics. Hence, retrieval systems for photographers are not necessarily harder to build. A key issue seems to be finding out the special needs of the photography world.

2.3 Arts

Can multimedia research help boost public interests in art appreciation and practice through the Internet?

Shih-Fu Chang: Not that I know.

Michael Lesk: I believe it can, but I would like to see us do something more advanced than the Hermitage-style QBIC example. There is too much reliance on color histograms. Some areas that might be worth looking at include:

1. Presentation of sculpture, either with inverse panoramas or 3-D modeling.

2. Stylistics research

Perhaps more important is that multimedia research also includes new creativity, and the Internet is often the way new kinds of art are distributed. I would like to believe that when artists create new work out of old (as with "sampling" in music) they will provide links to the source material, so that old as well as new work will be more accessible to the public.

Rainer Lienhart: NO. Multimedia is a new medium artist can use and many interesting artifacts will be created. It gives many new possibilities to artists. However, it will not transform people agnostic to arts to arts freaks.

Jiebo Luo: That is for sure.

The example I have is not for visual art but closely related. My young daughter just recently became hooked with American Idol. I found myself become a heavy user of YouTube after each episode, searching for music videos of the same song performed by the masters because she became very interested in how the renowned artists sang the same songs. It is amazing how much you can find in YouTube and you become so appreciative of the masters and the different styles, and ultimately become better at art appreciation. From an art appreciation perspective, the Internet in general and YouTube in particular in this case are influential. Granted, everything in YouTube is text search based.

Now back to (visual) art appreciation because we all work with images. My daughter happens to be into art as well. We do not have the equivalent of YouTube yet, but that may be something some of us can get rich of. Let's say I upload my daughter's own artwork to a Website, a multimedia search engine quickly extract features and recognize the style and subject of the artwork. It then provides links to either masterpieces that are related or online art lessons for the same type of artwork, it would thus help boost public interests in art appreciation and practice. It could even work as a B2C service in the same way as Google, i.e., link to museums, galleries, art studios and art schools.

Arnold Smeulders: As appreciation is concerned, it already did. Through the Internet, the public appreciation of museums has increased tremendously. As practice is concerned, artists have always used the forefront of technology and culture to express themselves. Whether it was the composition of Renaissance shifting from theocentric to civilians, the new dyes in paint, the new lubricants in paint, the advent of photography, the advent of motion pictures, the advent of video for the new art form of performance, and so on. So, my answer here is yes, and we do not have something to do with it. It will happen anyway.

Jia Li: It is desirable to have access to more digitized art works from different cultures. Currently, Western art is much more accessible through the Internet than other cultures. Art is a window to history and culture, and, in comparison with literature, requires little special training to understand. Through art, we can better appreciate different cultures and recognize common values. The barrier to broaden the presence of art in the digital world may not simply be technical. One obvious issue is copyright. Moreover, digitization is costly, and some museums may not appreciate its value. One starting point may be through the engagement of current artists around the globe. Sizable collections may be established from the contribution of these artists' own work, assuming they are well connected on the Internet. Software can be developed to compare art styles across cultures and over the history. Interesting links between art works may be discovered, and hence help artists find most inspiring or liked genres, possibly in a different culture. For instance, the traditional wax printing on cloths by natives of Yunnan province in China has the strong poster looking quality (flat, expressive, and bright colored) which some modern artists in the West have been attracted to and pursued.

As mentioned previously, the Internet is a relatively encouraging environment for people to show their art works even if they are not professionals. Practice of art either as a profession or a hobby can be helped by the Internet and multimedia research in several ways. Online communities may be built to provide critiques on art pieces. There are principles in making art. On the other hand, there is often no right or wrong in an artist's decision. The most touching part of an art piece is usually something personal to the artist, e.g., his/her understandings, feelings, emotions, or personalities. In a sense, there are no purely objective rules to guide an artist, and hence peer critiques form a powerful mechanism for improvement.

Similarly as with photography, one can improve skills of drawing and painting by looking at the works of others, especially those related in subjects or styles. This may boil down to the basic topic of content-based image retrieval. However, needs specific to art images should be addressed.

For instance, the vast majority of ancient Chinese ink paintings fall into a rather small number of genres in terms of what is depicted, namely, "mountain and water", "human figure", and "flower and birds" (more accurately, plants and animals). It is argued that, when retrieving these ink paintings, it is more meaningful to compare stroke styles, which are nontrivial to extract and characterize [3]. Similar scenarios arise in the study of oil paintings. For instance, art historians examine stroke characteristics (along with other measures) to authenticate paintings claimed to be by Van Gogh. To meet the new demands of art image comparison, researchers may have to explore at the fundamental level of creating new image analysis techniques.

2.4 Online Video Sharing

Can video or image retrieval better help people search online video sharing sites?

Shih-Fu Chang: Yes. This is more so than the case of helping content owners. When searching other people's content, we do not have the first-person information about the original context or events of the videos. Therefore, annotations like dates and events may not be meaningful and certainly not sufficient. In this case, any additional tags (like location, object, people) will be useful. Researchers working on content analysis and automatic labeling have a great chance to make contributions here.

Michael Lesk: Well, yes, this is one of the major research areas and has been for a long time. I think we need additional research on motion classification, to go with existing research on single images. Again perhaps the easiest way to serve needs is to recognize large classes of activity - not the very difficult problem of writing Labanotation from ballet video, but just recognizing that a video shows dancing.

Rainer Lienhart: There is no universal answer to this question, since I do not believe that video search is meaningful to all people. Do not expect it to be ever as omnipresent and important as the mobile phone or the Internet Search based on text. So some extended video search for the ordinary people is a little bit like the video phone. It shows a lot of promise and you can come up with nice scenarios, but it will never really take off. Only, in narrow usage scenarios, video search will be important. Therefore let me go through a few user groups:

1. Parents (like me): I am not a user of video sharing sites. I do not even know what to search. So how can a better search help me? It cannot. The only thing I can imagine is that I want to cross-correlate my video recordings with my photos, thus allowing better search for images/photos of my family members.

2. Kids: They will use the video search to find "cool" and/or illegal clips, or make sure that their video clips show up high on the video search. Most likely they will use it extensively as they feel that this is another place no parent can control.

3. Education: Video search might be important to better understand concept in biology, nature, science, engineering.

4. Surveillance: They have a huge demand for video search. Specialized systems will be developed to really accelerate the analysis of a crime or crisis. Or—in an optimistic view—it might even help to detect the preparation of a crime.

5. Entertainment: Well, text is all you need. I do not search for a main actor by showing the system a picture. This does not seem to be a common usage scenario. On rare occasions people will need it.

Jiebo Luo: Let me again use searching music video in YouTube as an example, which by the way as of early 2007 has 60% of the online video market with MySpace a distant second at 16%. It worked greatly in my earlier examples, but what if you do not know the title or artist name? For music, you love to search by hamming or something of that nature. For images and video, tagging is what is currently being used in Flickr and YouTube. Ultimately, we need content-based image and video retrieval (CIVR).

Given the difficulty of the general CIVR problem, I think a divide and conquer approach may work out. We all know the more constraints and domain knowledge you have, the more likely you can find a robust solution. What I meant is that we can build specialized CIVR systems, say for music video, sports video, news video, and so on, respectively. With a little bit of help from the user, for example, instructing the system that he is looking for sports video, the specialized CIVR engine can do a better job. If more automation is desired, we can build classifiers to recognize different content types (but allow the user to correct the system error).

I want to emphasize the fact that we do not have to have fully automatic systems. We just need to provide user-friendly systems. Users usually do not mind providing some minimum input or feedback. Designing a human-in-the-loop system that is enjoyable and engaging will go a long way. Remember, Google is not a perfect search engine but it is highly successful.

Arnold Smeulders: Yes, of course. Video retrieval for some questions is already at a level where text was in 1995 for all questions. So for those questions video search technology is already helpful. The difficulty is getting this differentiation in performance across to the public before their frustration runs high.

Jia Li: For sure, video/image retrieval and annotation techniques will continue to contribute to video searching and sharing. It seems that current commercial video search relies mostly on text. It would be interesting to see how video retrieval and annotation techniques can be incorporated into real-world large scale systems and help searching. Another interesting problem is to characterize and search based on composition ideas rather than the semantics of video clips. For instance, I saw a video clip in YouTube where every shot features the same guy dancing at a different place around the globe. The dance by itself is very casual. A sequence of shots showing different places alone is not interesting either. The combination of the two, however, strikes a deep impression. The world is so different from place to place, while so alike as well with the same dance greatly enjoyed everywhere. Creative composition approaches will add artistic flavors to home made videos. It is indeed difficult and maybe too abstract to describe composition ideas not to mention characterizing them by computers.

2.5 Social Networks

Are there new technical demands for multimedia in Internet social networks?

Shih-Fu Chang: This is the most exciting area to work on now. There are great opportunities to apply data mining techniques to explore the best use of the rich media content and the dynamic linking structures in social networks. There are a lot of interesting questions to ask. How do people tag and link their photos when they upload them? How do people comment and cross-reference photos of their friends and others? How do we explore such structured and dynamic information to increase the searchability of the content?

Michael Lesk: I am afraid this is more driven by law than technology. Right now the social networks (YouTube) have found that people like clipping commercial entertainment, just as I might put a cartoon torn from a magazine on my office door. But using a cartoon that way is legal, and posting a clip from almost any TV program or movie is not. So I think we need procedures for very easy animation, so that amateurs can make their own videos to post when they want to go beyond making a video of their cat. We could use something in the way of "citation" for entertainment video, so that people could refer in a legal way to a scene that, one hopes, could then be found on a library/studio site.

Rainer Lienhart: I have a family - I do not need the Internet to have a social network (ok, that statement is over painted, but I guess there is some truth in it). This question should be answered by someone who is engaged in social networks.

Jiebo Luo: Several points come to mind.

1. *Employ multimedia search to build social networks*
   Link people through recognition of places, people, subjects, and events. It will be enabled by some form or combination of semantic tagging (automatic, semi-automatic, or manual). Again, it does not have to be automatic so long as the interface is user friendly.

2. *Enable user generated content (UGC) by multimedia search*
   People will be able to take their raw photo and video collections, turn them into more creative and more entertaining asset to share, and thus open the doors and channels for more social connections. In this regard, it may be imperative to provide iTune type of asset repository so a user can blend his content with copyrighted content in a legitimate way.

3. *Multimedia search and management on a mobile device*
   More and more people want to communicate on the go. More and more devices are invented to respond to that need (e.g., iPhone). Granted, the newer devices are equipped with greater computing power, better display, higher storage, and longer-lasting battery. However, many of the content recognition algorithms are still computationally expensive, and one cannot store everything he needs on the mobile device. So a client-server type or distributed computing type of infrastructure may be critical for multimedia search and retrieval, i.e., using networked resource to perform multimedia search and management on a mobile device.

4. *Personalization*
   Yes, everyone will be immersed in the social networks. However, the multimedia search and management tools need to be personalized for higher efficiency and higher satisfaction. This will pose challenges and opportunities for multimedia research (e.g., incremental learning, individual learning, user adaptation).

I would like to end by quoting the photography is a universal language so digital photography (in a broad sense to include photos and videos) will undoubtedly facilitate social networks for the increasingly connected and fast growing digital society, crossing even culture and nation boundaries.

Arnold Smeulders: Most prominent is the handling and verification of socially provided information, intrinsically uncertain, conflicting, and (sometimes deliberately) false as it may be.

Jia Li: Conceptually, I view this issue as the interaction among (1) social networks, (2) semantic and visual networks, and (3) multimedia technologies (Figure 1).

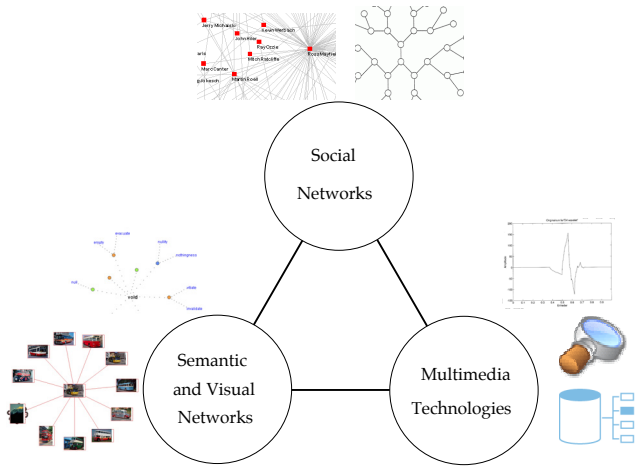

Figure 1: The interaction among social networks, semantic and visual networks, and multimedia technologies.

In the past, when we were developing multimedia technologies, we made use of techniques such as relational databases, signal and image processing, and machine learning. However, we have rarely leveraged the massive amount of multimedia objects or the social networks associated with them. With the growing availability of such data sets, we can potentially integrate information from a wide range of sources. For example, with the WordNet semantic network and pairwise image similarity within a large image collection, we can find and use visually representative images to illustrate a story. When we have a network of users, we can build a multimedia retrieval system that bridges inaccurate machine-generated tags with incomplete human-supplied tags. The technical challenge, of course, is to efficiently leverage massive amount of information, often of different formats and structure, in the system.

On the other hand, computational technologies are needed to help build social networks, connecting people on the network using multimedia visual networks and semantic networks. Up to now, most online social networks have been large depositories of shared multimedia objects. They do not actively assist users find new friends. Nor do they

assist users in making use of the network beyond sharing of contents. Technologies can be developed to do much more. For instance, we can help users find other users with similar artistic tastes or interests using a querying by example retrieval system. Computer systems can also rate the contribution of users based on the visual quality of their uploaded multimedia objects, as compared against user-rated objects on the same network. Users with high ratings can then be promoted as leaders in the network.

3. CONCLUSIONS

The panelists discussed several topics in multimedia research that are of high potential importance with the rapid growth of the digital society. These topics are (1) personal multimedia management, sharing, and archive, (2) whether multimedia research can help people enhance photography skills, (3) help boost public interests in art appreciation and practice, (4) enhance online video sharing, (5) and offer technical assistance to Internet social networks.

4. ACKNOWLEDGMENTS

The research of Shih-Fu Chang, Michael Lesk, and Jia Li has been funded by the US National Science Foundation. The research of Rainer Lienhart and Arnold Smeulders has been funded by the European Commission. Jiebo Luo is funded by the Kodak Research. Jia Li is also affiliated with Google Research Lab in 2007-2008. The opinions expressed in this article are their personal views. Alberto Del Bimbo and James Z. Wang assisted in developing this panel.

5. REFERENCES

[1] R. Datta, D. Joshi, J. Li and J. Z. Wang, "Studying aesthetics in photographic images using a computational approach," *Proc. the European Conference on Computer Vision*, Part III, Lecture Notes in Computer Science, vol. 3953, pp. 288-301, Graz, Austria, May 2006.

[2] D. Joshi, J. Z. Wang and J. Li "The story picturing engine - a system for automatic text illustration," *ACM Transactions on Multimedia Computing, Communications and Applications*, vol. 2, no. 1, pp. 68-89, 2006.

[3] J. Li and J. Z. Wang, "Studying digital imagery of ancient paintings by mixtures of stochastic Models," *IEEE Transactions on Image Processing*, vol. 13, no. 3, pp. 340-353, 2004.

[4] J. Luo, M. Boutell, C. Brown, "Pictures are not taken in a vacuum - an overview of exploiting context for semantic scene content understanding", *IEEE Signal Processing Magazine*, vol. 23, no. 2, pp. 101-114, March 2006.