Visual Signal Processing and Communication Technical Committee Vision Statement

The mission of Visual Signal Processing and Communication (VSPC) Technical Committee (TC) at IEEE Circuits and Systems Society is to promote research and development activities and advance state of the art in visual signal processing and communication with respect to circuits and systems. On the one hand, it leverages new development in circuits and systems to advance the visual signal processing and communication field; on the other hand, and more importantly, it studies the fundamental problems in visual signal processing and communication, develops enabling technologies for emerging applications related to the concerned field, and influences the directions of research and development in circuits and systems. More specifically, VSPC TC addresses the core component technologies at every stage in the lifecycle of visual signals that is related to signal processing and communication, including acquisition, processing, coding, communication and networking, post-processing, synthesis and rendering, presentation and display, analysis, recognition, understanding, retrieval, search, security, management, monetization, architecture and implementation. While applications and systems are not a focus of the VSPC TC, it will also address core technologies aiming for closely coupled components as the ultimate performance of a system often comes from joint optimization of a number of related components.

Recently, we have witnessed several important trends in theory and applications of circuits and systems, information technology, computer science, and the Internet. We believe these new trends will bring profound impact to the visual signal processing and communication field and imply many important directions for research and development in the next 5-10 years.

1. **Impending breakthroughs in information theory.** With the new development of compressive sensing theory, machine learning theory, and other related information theory, we envision a complete paradigm shift in visual signal processing and communication. Signal acquisition may depart from the traditional Nyquist-Shannon sampling theorem. Instead it may be based on compressive sensing theory that exploits the intrinsic highly-correlated property in natural visual signals. Many traditional visual processing and communication pipelines need to be studied and changed to embrace this new trend. Visual signal processing and communication may also shift away from pixel-wise processing in the traditional Hilbert space to structure, region or object based processing in a new metric space. Advanced development in machine learning may lead to a mass-knowledge based visual signal processing and communication system instead of traditional memory-less or memory-based signal processing and communication pipelines. New theory on distributed coding and distributed processing is also emerging to enable independent compression and processing of visual signals with much reduced communication overhead but without degradation of overall performance. Shannon's information theory may be updated to reflect the changes in

modern visual signal processing and communication that he didn't foresee. For example, if we redefine information as the new knowledge perceived by the user, a web search operation might be readily formulated as a filtering process.

2. **Vast amount of web data with embedded knowledge.** The Internet is a massive computing platform with vast amount of web data containing more and more human knowledge. It has the largest collection of text documents, photos, audio data and video data ever in human history with lots of the visual data annotated with tags or linked with related text data. Though noisy, it may provide the most comprehensive knowledge base ever for visual signal analysis, recognition, understanding, retrieval, search and management. It could also provide a new framework for knowledge-based visual signal compression and processing which could lead to new performance breakthrough in these areas. The key research problem here is how to extract and represent knowledge related to visual signals in the presence of noise, how to leverage this kind of knowledge in real world visual signal processing and communication and how to feedback the processing results to further enhance the web knowledge.

3. **Human in the loop.** The reason for inventing automatic machines is to liberate human from tedious labor. Though computer has been invented for several decades but it still could not handle some very simple task with confidence, for example, object recognition and content understanding for visual signals. Many of these tasks can be completed simply by leveraging human intuition or intelligence. Aforementioned massive scale web data does bring us a lot of human knowledge. However, it is based on collective human intelligence of many years' accumulation. There are others areas that human can help in visual signal processing and communication. Collaborative filtering of visual contents is based on the behavior of a group of people with similar interests. Large amount of human annotated visual signal, be it as photos or videos, can provide ground truth for the machine learning engine in visual signal understanding and recognition. Social networks on the other hand can create viral communication/distribution of visual data (for example, P2P communication); motivate people to tag the visual data; rate and comment on the quality of the visual data; and promote user generated contents (so called social media). By bringing human in the loop, collective human intelligence, user usage data and web data could provide many out of band side information to make the visual signal processing and understanding much easier.

4. **Wide adoption of cloud computing.** The Cloud Computing age is coming, regardless if you are ready for it or not. Besides advantages of a more cost-effective and efficient computing platform, it also provides instant access to vast amount on-demand computing resources and easy access to the massive web data. As a platform, it allows an always-available knowledge base sitting in the cloud and offloading many visual processing tasks to the cloud. This is especially true for those thin client devices. With the help from the cloud, it makes many traditionally impossible visual processing tasks possible, including many content understanding and object recognition tasks. The core problems are how to partition visual processing tasks between cloud and client, how to solve the QoS and adaptation problem of visual signals in the cloud, how to scale simple client-side visual signal processing algorithms to handle web-scale data using the cloud computing platform, and how to protect the visual

data from security breach and privacy invasion while not degrading the processing performance.

5. **Continuing advancement of rich client.** While cloud computing will mark a new era in computing history, the client device continues to evolve to a much more powerful platform. On the one hand, highly parallel homogenous processors such as multi-core and many-core CPUs and GPUs are becoming commodity. On the other hand, multiple heterogeneous devices are more and more readily at hand. Visual signal processing and communication algorithms should be designed with both homogenously and heterogeneously parallel processors (including cloud processors) in mind to take advantage of such great advancement in rich client. Moreover, the increase in the varieties of clients also poses challenges in device adaptation that enables the best user experience seamlessly on any devices.

6. **Ubiquitous and rich varieties of commodity sensors – Internet of Things.** Sensors and sensor networks are becoming rich in varieties, ubiquitous and more and more affordable. First, new sensors may generate many 2D or 3D signals that are not visual signals in traditional sense, but they can be processed by similar visual processing technologies, for example, multi-touch sensors, depth sensors, infrared sensors, and motion sensors, etc. Processing of these new kinds of "visual" signals will enable new functions such as new human computer interface such as gesture and touch, etc. Second, new sensors can narrow down the context for visual signal processing or directly provide out of band side information to facilitate visual signal processing, for example, a GPS sensor can provide a very precise context to help recognize a landmark image; a depth sensor can greatly help to extract the contour of an object; and a motion sensor might be able to detect which parts in a video are moving. Third, visual sensors are also moving from 2D sensors to 3D sensors, from a single sensor to an array of multi-view sensors, or even randomly placed visual sensor networks. These new sensor types will enable new user experiences, such as 3D video, multi-view video and surveillance video networks. It will also bring new challenges in visual signal processing and communication research, for example, 3-D and multi-view coding, distributed visual coding and object tracking across visual sensors networks, tradeoff of communication cost and processing cost on individual visual sensors.

7. **Ready for mobile and wireless devices.** Nowadays, it is hard to imagine any visual signal processing and communication technologies that cannot work with mobile and wireless devices. By leveraging the new development in rich client and cloud computing, sophisticated visual signal processing and communication tasks can be enabled. It is very important to jointly optimize the system performance with regard to available bandwidth, processing power, battery life, delay, display, user-perceived experience. Ideally, visual data, processing software, and screen rendering can all be flexibly placed to either in the cloud or on mobile devices dynamically adaptive to the current context. Visual signal processing and communication technologies that address the asymmetric nature of uplink and downlink in wireless communication; and the offline experience through smart caching would be much needed for mobile and wireless devices. A long existing problem on mobile devices is the limitation of display size. With the development of 3D display, paper-like rollup display and

pico-projector based display, larger and more vivid display is becoming a reality. It might need new technologies to address visual signal processing and communication issues for such a large or 3D display. Another important area is new communication scheme that jointly considers source coding, channel coding and even modulation for wireless communication based on visual signal characteristics.

8. **Blurred boundary of physical and virtual world – Augmented Reality.** Pursuing natural user interface is a constant goal for computer engineers and a dream for consumers. With the advance in visual signal processing and communication, visual signal synthesis and computer graphics, display technology, sensor technology and location-based services, etc, blending the physical world with virtual world as a new form of natural user interface is becoming a reality. From visual signal processing and communication perspective, a lot of challenges are involved, for example, precisely locating the position of a visual signal in 3D space (both global position and local position and orientation), recognizing the objects in the visual signal, and seamlessly synthesizing the corresponding virtual objects into the physical world captured by the camera. Augmented reality represents a new generation of nature user interface and is worth a lot of attention and investment from the VSPC community.

9. **Embracing green Computing.** Green computing has quite a few implications to visual signal processing and communications. The most significant one should be energy saving. Low power consumption should be an important design constraint whether it is for algorithm design or for real implementation. From system level, it means we should only generate and deliver the results that exactly match the user's expectation and within the device's limits. For example, delivering a HD content to a mobile device that can only display SD resolution is an overkill that wastes both bandwidth and decoding processing power; recognizing more objects in a visual signal than necessary to meet the particular user's interest is also a waste of computing power; retrieving visual signals that are not in the current context maybe a waste of cloud computing power, etc. Of course, reusing existing components to realize a certain visual signal processing and communication function is another aspect towards green computing.

10. **Integrated system level research.** From the above discussions on the various trends and impacts to visual signal processing and communication field, it can be observed that one of the most important trend in visual signal processing and communication is that it starts to move away from traditional independent component-wise research to more system level research. While there is still a need to perform component research for each stage in the visual signal lifecycle, it should be done at system level in order to deliver the best possible results. A few key words should be included into future visual signal processing and communication research, that is, knowledge-based, context-based, collective human intelligence, cross-component optimization, adaptation and scalability, distributed and parallel, natural, and green. Specifically, for the scalability issue, it involves many aspects, including dataset scalability, class/category number scalability, dimension scalability (from 2D to 3D to multi-view), source coding scalability, channel coding scalability, network scalability, device scalability, user scalability and adaptation, scalability between cloud and

client, etc.

With the above trends, the Visual Signal Processing and Communication Technical Committee would focus on the following activities for the next 5-10 years with consideration of the implications from these trends.

- **Information Theory:** New information theory such as compressive sensing and its impact to related components in visual signal processing and communication.
- **Visual Coding:** New visual coding schemes, including next generation international standards, 3D video, and 2D and 3D graphics, scalable coding, compression of visual features for content analysis, and multiple description coding.
- **Visual Content Analysis**: Visual content processing, analysis and understanding based on web data, human contributed data, sensor data, and context data, including new applications such as computational photography, semantic video representation, affective computing, visual quality assessment and quality of experience assessment and management.
- **Communication and Networking**: including joint source, channel and modulation optimization; network coding; QoS for visual signal communication in could computing; etc.
- **Visual Sensors:** Acquisition, processing and communication of 3D video, multi-view video and video from visual sensor networks.
- **NUI:** Natural user interface based on visual signal processing techniques including augmented reality, multi-touch; gesture-based UI.
- **Security and Privacy Protection**: Security and privacy in cloud computing for visual signals, forensics and steganalysis, encrypted domain processing.
- **Parallel and Distributed Processing**: Highly parallel visual signal processing algorithms and implementation, including visual coding and processing over multi-core and many-core CPU, GPU and dedicated ASICs; distributed visual signal processing and coding.

As predicting future is always risky, the VSPC TC should continuously refine this document when new information is available.

Appendix E: Excerpt of VSPC TC Members' Inputs

Enrico Magli

One emerging area I am working in is "compressed sensing". This is very important for SPS, but will likely become important in CAS in the coming years. Please find below a paragraph on this topic. It is not necessarily focused on anything specific, but just a description of the topic.

Existing signal acquisition schemes are based on Shannon's theorem, which requires the sampling frequency to be at least twice as large as the signals maximum frequency. Compressive sampling (CS) is a new signal acquisition and processing paradigm invented in 2005, which revolutionizes Shannon`s sampling by exploiting the notion that most natural signals, e.g. images, are highly correlated. Correlation implies that there exists a domain in which the signal is sparse; for example, only a small fraction of the wavelet coefficients of natural images are significantly different from zero. CS investigates the problem of acquiring a set of measurements comparatively much smaller than dictated by Shannon`s theorem, from which the signal can be reconstructed exactly or almost exactly. The main objective of CS is not to perform compression; rather, CS aims at avoiding altogether the acquisition of a very large number of samples, thereby allowing to design sensors that are more effective at acquiring the signal of interest. Interestingly, CS can be carried out using surprisingly simple techniques. A "measurement" is obtained as the scalar product of the signal with a pseudorandom sequence. The reconstruction of the signal requires to find the sparsest signal (or transform thereof) that matches the available measurements, which can be performed using, amongst others, linear programming techniques.

C. C. Jay Kuo

In the area of visual signal processing, I expect some major breakthrough in automatic visual content understanding, summarization, indexing and retrieval due to the rapid development of visual representation and machine learning techniques in recent years. Another quick growing field is 3D video, including acquisition, compression, transmission and processing.

In the area of visual signal communication, people are developing a new generation video coding standard targeting at high definition video. The standardization effort is expected to last for 2-3 years. After that, people will focus on its performance optimization and real world applications. The combination of these two will be the main activities in the next decade.

For both computer vision and multimedia, I think the scalability (class/category number, data sample number, and dimension from 2D to 3D) shall be the main focus.

Yun-Qing Shi

In ISCAS10, there is a short discussion between Wenwu and me. That is, security will be one of the key issues in Cloud computing. Hence, this may be an issue for this TC to address.
Apparently, people are talking about Cloud Computation more and more these days. It seems cloud computation becomes the next generation of computational framework. The spirit of cloud

computation is distributed computation in a revolutionized larger scale, as compared with before. With large amount of critically important data being transmitted swiftly and remotely, security and privacy turn out to be critical issues. For instance, the access of sensitive data and computation becomes a serious issue. Although security and privacy have been studied for years, these issues remain challenging. With much wide scope of cloud computation, much more efforts are called for. From circuits and system point of view, the VCSP TV will and can make significant contribution in the research on security and privacy for cloud computation.

Wenwu Zhu

I agree with you (Yun-Qing), multimedia security and privacy is foreseen to become important issue in Visual/multimedia Cloud Computing, such as video/photo editing, video/photo sharing, video/image rendering, video streaming, etc. When personal visual data are being moved to public cloud or even private cloud for visual computing/processing, privacy and multimedia security is foreseen to become an important issue, as you just pointed out.

Jiangtao (Gene) Wen

Yet another area of potential interests in the future is computational photography. With the advance of CMOS imager technology, it is possible now to control the exposure of individual pixels during acquisition of video. Cameras integrating gyroscopes and various sensors as well as GPS will also open new possibilities for real time on-the-spot processing that was never possible before. I believe this will become an important topic in the coming 5 years.

Tony Ho

Again fully agree and support the areas that Alex and Yun have proposed on looking into systems level research and implementation of multimedia security, forensics and steganalysis techniques.

These topics will complement very well with the on-going theoretical work and analysis in multimedia security and related areas.

Weisi Lin

Perception-inspired visual modeling and processing can be a potential area of more active research, because humans are the ultimate consumer for most (if not all) products and services; this also adds a new dimension to the improvement space of video and image manipulations. In addition, truly integrated and effective audiovisual modeling can be a good direction since visual signal seldom exists alone. The data-driven approach (e.g., via using the data available in the Web) may open new opportunities for multimedia processing.

Frossard Pascal

I think that distributed processing, vision sensor networks, new sensing modes and 3D are probably some of the keywords we should have, if they have not been proposed already.

Ling Guan

We may consider adding affective computing which will heavily influence the next generation multimedia research and technology.

Jian Zhang

Video content representation and understanding

The aim of this research is to achieve semantic and structural representation of video content to enable meaningful content search and retrieval. Topics include: video summarisation and table of content generation, automatic and semi-automatic annotation of image/video, and semantic video representation. Using web based image/video data is a new direction in ACM MM and CVPR.

Low Level Feature Extraction and Content analysis

It provides content features and object information to enable meaningful regional based image and video content representation. Technologies include low level feature extraction, clustering and feature word generation. A key challenge is to research robust and low complexity algorithms for the need of real-time systems.

Video coding and communication

Video coding provides enabling technology for video archiving, adaptation and transmission to a variety of networked receivers. One challenge is to transmit image/video data over heterogeneous networks such as WLANs and UWB. Challenges also exist in video adaptation to deliver personalized video content to users with different terminals and usage scenarios.

Ngan King Ngi

I propose to include in the TC vision plan the field of "Visual Quality Assessment of Images and Videos".

Visual quality assessment of images and videos has become important nowadays with the demand of high-definition and high-quality viewing experience. The standards bodies, i.e., ITU-T VCEG and MPEG, now require subjective tests to be performed on all proposals. Research topics include modeling of human perceptual mechanism and development of with-reference, near-reference and no-reference objective metrics of subjective quality for images, videos and 3DTV.

Jian Fei Cai

3D video processing & networking: we can vision that media content is going to be much richer (such as 3D media) and more diverse (mixed with different types of media), and the way of consuming media is going to be more immersive (beyond anywhere, any time) . All these require more profound processing and networking techniques. In particular, network processing and end-point signal processing might need to be integrated in a tighter way to eventually provide users real immersive experience.

Chen Zhi Bo

My feedbacks:

1. Media Quality of Experience assessment and management: including media analysis, human vision system, metadata extraction, cross-layer optimization, etc.

2. 3D Computer Graphic content representation, coding and distribution

Ebroul Izquierdo

Multimedia information retrieval: Automated multimedia structuring, tagging, low-level and semantic-based image/video/audio annotation. Access and retrieval of multimedia data.

Scalable video source and channel coding: Including aspects related to Multiple Description Coding and Network Coding.

Shuicheng Yan

For both computer vision and multimedia, I think the scalability (class/category number, data sample number, and dimension from 2D to 3D) shall be the main focus.

Feng Wu

(1) Highly parallel image and video coding algorithms for multi-core and GPU. (The reason to propose this topic is obvious. It has become much cheaper to own more than 32-core CPU and highly parallel processing GPU. However, image and video coding is still to process block by block. It is desired to develop new algorithms for highly parallel image and video coding.)

(2) New hardware architecture and design for highly parallel image and video coding. (It is related to the first topic. But it focuses on hardware part because this TC is in CASS.)

Oscar Au

I definitely agree with Feng that an important topic is the highly parallel image and video image and processing algorithms for multicore and GOP. Relevant hardware architecture will be important too.

Another obvious topic is video coding. MPEG and ITU-T VCEG just agreed to jointly develop the next generation video coding standard. If history is an indication, it will very likely to be moving fast and very important. With lots of potential monetary implication, it will be a hot topic for the next few years.

Cloud computing is very hot. However, I believe people are still exploring its implication on circuits and systems and the associated signal processing problems. Like what Alex, Yun and Anthony said, data security/hiding/forensic will be important topics. I believe that the highly parallel algorithms for heterogeneous systems will be important. Encrypted domain process will be important too.

Green circuits and systems are likely to be hot also. I just attended a Green computing conference in Shanghai (ICGCS) organized by DSP TC. The topic will attract a lot of attention.

Shiqiang Yang

I think should consider to add the social media as a topic which will influence the multimedia and network in the future.