IEEE MULTIMEDIA COMMUNICATIONS TECHNICAL COMMITTEE

E-LETTER

Vol. 4, No. 6, July 2009

CONTENTS

IEEE COMMUNICATIONS SOCIETY

Message from Editor-in-Chief

It is very nice to greet everybody again after coming back from the beautiful Dresden and the ICC 2009. The MMTC had a wonderful meeting during the conference and we have seen many old friends as well as new faces in the meeting. Drs. Qian Zhang and Bin Wei have come up a wonderful meeting minutes to share all the details of the meeting to our fellow members including a few photos taken there. Please read the report at the pages 3-5.

In this Issue, we feature a Special Issue on *Visual Analysis for Communications*, which contains six position papers from world top scientists in the field. In addition, in the Editor's Selected Book Recommendation column, our Column Editor, Dr. Shiguo Lian (France Telecom R&D, China) recommend the book "Wireless Multimedia Communications: Convergence, DSP, QoS and Security" by K. R. Rao et al. Please check out more details in the article.

Furthermore, in the focused technology column, Dr. Deepa Kunda (Texas A&M University, USA) highlights the challenges in the emerging multimedia sensor networks (MMSN), and introduces the concept of employing directional link networks to address these difficulties in the MMSN system. At last, in the IG column, Drs. Madjid Merabti (Liverpool John Moores University, UK) and Heather Yu (Huawei Technology, USA) give a brief introduction on the home networking interest group with the associated research topics and the exciting events that have been run by this IG.

Now let us explore more inside the Special Issue on *Visual Analysis for Communications*. The papers can be classified into three categories:

(1) Object segmentation and tracking

The first three papers fall into this area. In the first paper delivered by Drs. K. N. Ngan (Chinese University of Hong Kong, China) and H. Li (Univ. of Electronic Science and Technology, China), the evolving history and key ideas in semantic object segmentation have been briefly reviewed and the challenges are highlighted; In the second paper, Prof. T. N. Pappas (Northwestern University, USA) demonstrates how to use mathematical and perceptual models for image and video segmentation with rich examples and results; After that, Drs. S. Lee (Dongseo University, Korea) and M. G. Kang (Yonsei Univeristy, Korea) demonstrate a centroid shift based object tracking algorithm as an alternative solution of mean-shift tracking, which has low computation costs and good stability.

(2) Large scale content management

The paper delivered by Prof. B. S. Manjunath (University of California, Santa Barbara) discusses the emerging challenges in large scale content management, such as meaningful video summary, activity analysis, content retrieval, and the distributed content processing capability.

(3) Video analysis for coding & communication

In the last two papers, the applications of video analysis for coding and communications are covered. The paper by Dr. P. L. Correia (Instituto Superior Tecnico, Portugal) discusses how the segmentation and feature extraction can be used to optimize the coding parameters in video communications. Drs. P. Nunes (Instituto Superior de Ciências do Trabalho e da Empresa, Portugal) and F. Pereira (Instituto Superior Técnico, Portugal) overview the rate control approaches adopted in the video coding standards, and point out major challenges.

Clearly there are many more challenges and issues that have not been fully covered in this Special Issue. Hopefully we will get them covered in our future Issues.

As always, I thank all Editors of the E-Letter, and our authors to make this issue successful.

Thank you very much.

Haohong Wang
Editor-in-Chief, MMTC E-Letter

HIGHLIGHT NEWS & INFORMATION

MMTC Meeting Minutes

IEEE ICC 2009, Dresden, Germany
Date: June 16, 2009
Time: 12:30-14:00
Room: Seminar #1, Convention Center
Organized by Committee Chair Qian Zhang, minutes by Bin Wei

Agenda

0. Informal discussion and networking time
1. Welcome new members / introduction
2. Last meeting minutes approval (Globecom 2008)
3. MMTC Best Paper Award 2009 winner announcement
4. Report on conference activities
    a. CCNC 2009 (John)
    b. ICC 2009 (Suba)
    c. ICME 2009 (Wenjun)
    d. Globecom 2009 (Qian)
    e. ICC 2010 (Zhu)
    f. Globecome 2010 (Bin, Haohong)
5. Discuss the issues related to ICME
6. GITC and TAC report (Heather)
7. MMTC IGs reports - all IG chairs
8. Sub-committee report (report for the recent changes)
9. Report for News Letter activities (Haohong)
10. Publication Report (e.g., activities in terms of special issues on IEEE journals/magazines)
11. Suggestions & discussions – everyone
12. Adjourn

1. Welcome new members and introduction

Meeting convened by the committee chair Qian Zhang. This is the second MMTC meeting after Qian Zhang is elected. The meeting started at 12:30 local time with 24 attendees. Around the table introductions were made.

2. Last meeting minutes approval (Globecom 2008)

Committee approves it.

3. MMTC Best Paper Award 2008 winner announcement

Award subcommittee - Chair: Dr. Dapeng Wu; Members: Prof. Oliver Wu (University of Florida), Prof. K.P.Subbalakshmi (Stevens Institute of Technology), Prof. Marco Roccetti (University of Bologna), Prof. Chang Wen Chen (University of Baffalo), Dr. Olivier Verscheure (IBM Research), Prof. Zhihai He (University of Missouri Columbia), and Prof Jianchuan Liu (Simon Fraser Univeriisity).

Two papers are selected for the best paper award:

1. Mea Wang, Baochun Li. "R2: Random Push with Random Network Coding in Live Peer-to-Peer Streaming,“ in IEEE Journal on Selected Areas in Communications, Special Issue on Advances in Peer-to-Peer Streaming Systems, Vol. 25, No. 9, pp. 1655-1666, December 2007.

2. B. Li, S.-S. Xie, G. Y. Keung, J.-C. Liu, I. Stoica, H. Zhang and X.-Y. Zhang, "An Empirical Study of the Coolstreaming+ System,“ IEEE Journal on Selected Areas in Communications, Special Issue on Advances in Peer-to-Peer Streaming System, 25(9):1627-1639, December 2007.

Dr. Mea Wang is present to receive the award at the announcement.

4. Report for the conference activities (mostly presented by Dr. Qian Zhang)

- ICC 2009:
    - Communications Software and Services Symposium. Representative: Giovanni Pau, 160 papers registered, active 114, 4 redirected to another symposium.

  - Symposium on Selected Areas in Communications. Representative: K.P.Subbalakshmi and M. Merabti, 125 papers submitted, about 36% accepted, 49 papers accepted, 5 tracks: Cognitive networks; Data storage; Vehicular communications; Situation management; Networked services.
- CCNC 2009: CCNC09 had a record number of over 450 full paper submissions, a 33% increase from the previous year. In addition, short paper submissions were 217, an increase of 100%. The final program represents a highly selective and high quality set of research results. Of the 335 papers submitted to the six technical tracks, only 117 were accepted, for a very competitive acceptance rate of 35%. Of the 122 papers submitted to special sessions, 61 were accepted. Of the 217 short paper submissions, only 62 were accepted. Of 52 demonstration proposals, 29 were accepted.
- GC 2009: Communications Software and Services Symposium, representative: Qian Zhang. Papers received 110, reviewed 105, accepted 36, acceptance ratio 34.3%, number of oral sessions 5 and number of poster sessions 1.
- ICC 2010: Change symposium title to include "multimedia communications", representative: Zhu Li
- GlobeCom 2010: title "communication software, services, and multimedia applications", representative: Bin Wei. This symposium is relative small compared with other Globecom symposiums, in terms of paper submissions. We really need our TC member's support to make this symposium successful by submitting more papers.
- ICME 2009 (Wenjun slides): MMTC is one of the sponsors and will still show our support. MMTC organizes a workshop at ICME 2009. Pre-steering committee is scheduled and we will have 3 representatives, Wenjun Zeng, Gary Chan, and Pascal Frossard to attend.

5. GITC and TAC report (Heather)

In ICC 2011, our symposium is missing. We need to make sure to contact the TPC chair about it.

6. MMTC IGs reports – all IG chairs

Cross Layer (Song); Mobile and and Wireless Multimedia (Mouli); Automatic Communications (ACIG).

7. Subcommittee reports

MMC TC established the following subcommittees:

Membership Development Subcommittee Chair: Chang Wen Chen;

Conferences Subcommittee Chair: Pascal Frossard;

Publications Subcommittee Chair: Gary Chan;

Awards Subcommittee Chair: Dapeng Wu;

Standards Subcommittee Chair: Stanley Moyer;

TC Promotion & Improvement Sub-Committee: Chair: Haohong Wang.

Dr. M. Merabti "what is supporting?" New may take away existing business. What is MM communications? We are a community.
Qian: there are rules about the roles from IEEE. We need to promote supports first and then consider how to make it under proper management.

Haohong talked about TCPI, the need to setup the IG. A few tasks: 1. Bring new members; 2. Bring some strategies to revive the members involvement; 3. Support more paper submissions; 4. Re-activate MMTC E-letters. Future development: bring friends to join MMTC; help us with E-Letter as an editor; Start to nominate new editor-in-chief in Sept.

8. E-Letter

Editor-in-chief: Haohong Wang, editors: Phillip Roose, Chaonggang Wang, Guanming Su, Shiguo Lian, and Antonios Argyriou. New issue was out on Feb 2009. Every month there has been one since then.

9. Suggestions & discussions – everyone

Enhance the impact of MMTC in key conferences: ICC/GlobeCom, CCNC and ICME; Need to be more active in seeking promotions of our members to "senior member" for our TC members and work with several distinguished member towards "fellow" elevations (Prof. Chang Wen Chen); Special issues with the endorsement of MMTC.

10. Adjoin

The committee chair, Dr. Zhang, thanks everyone's effort for joining the meeting. We would also like to thank her for organizing this meeting and providing the information for our MMTC members.

11. Attendee List

Heather Yu	Huawei Technologies, USA
Qian Zhang	HKUST
Martin Reisslein	Arizona State University
Nirwan Ansari	NJIT
Haohong Wang	TCL-Thomson
Mea Wang	University of Calgary
Jiming Chen	Zhejiang University
Maja Mati, JSEVIC	University of Zagres
Iztok Humar	University of Ljubcjana
Christian Hoene	University of Tubingen
Shiguo Lian	Orange Labs Beijing
R. Chandramouli (mouli)	Stevens Institute of Tech.
Saverio Niccolini	NEC Laboratories Europe
Christian Oslleasri	University of Pisa
David Iacono	University of Pisa
Ivan Bajjc	Simon Fraser University
Larry Xue	Arizona State University
Madjid Merabti	Liverpool More University, UK
Byrav Ramamurthy	Univ. Nebraska-Lincoln
Raj Jain	Washington University in Saint Louis
Min Chen	University of British Columbia
K.P.Subalaksh (Suba)	Stevens Institute of Technology
Joel Hodrigues	University Beira Interior, PT
Bin Wei	AT&T Labs – Research

TECHNOLOGY ADVANCES

Distinguished Position Paper Series

Semantic Object Segmentation

King N. Ngan (IEEE Fellow), The Chinese University of Hong Kong, China
Hongliang Li, University of Electronic Science and Technology China, China
knngan@ee.cuhk.edu.hk

We often say "a picture is worth a thousand words". This is because the semantic content of a picture cannot be easily described by a few words. Have you ever wondered when you look at a picture, how do your eyes and brain extract information from the scene? What is the basic unit for the image understanding when a person observes an image or video scene? The possible answer may lie in the semantic objects, which can provide the meaningful cues for finding the scene content. However, in order to understand the semantics of the scene content, we need to separate or segment the content into its constituent parts. This has posed many challenges to vision researchers over the past decade and so far arbitrary object segmentation from images or videos is still an unsolved problem.

The extraction of semantic objects plays an important role in digital image/video processing, pattern recognition, and computer vision. It provides an efficient way to bridge the primary image data and semantic (and thus meaningful) content in image understanding. The task of segmenting a semantic object can be found in many fields, such as video surveillance, content-based video summarization, content-based coding application, computer vision, video conferencing, and digital entertainment [1]. A semantic object can be exploited to provide the user with the flexibility of content-based access and manipulation such as fast indexing from video databases and efficient coding of regions of interest [2]. An interesting example can be illustrated by the cartoon face in real-time video, where the extracted face region is used to construct the basic face contour, and helps to perform the animation of face expression.

Earlier techniques of image segmentation can be traced back over forty years, which aimed to extract some uniform and homogeneous regions based on the edge, color, or other low-level features. These methods are usually called non-semantic modes. Among them, a boundary-based method is mostly used in segmenting image regions, which first employs the gradient operator to extract the edges, and then groups them into object contour. The main drawback of this method is the lack of robustness during the contour closure due to the difficult computation of the region's boundaries.

In order to satisfy the future content-based multimedia analysis, more and more researchers seek for the semantic object segmentation by discovering the meaningful regions. However, this method usually needs to first identify the position of semantic objects from images by performing the spatial classification based on a certain prior knowledge. An intrinsic problem of this progress is that there is no unified method that is available for detecting all semantic objects. To avoid the complicated object recognition process, an interactive image segmentation approach based on graph-cut optimization has been developed which extracts a semantic object at the cost of interactive effort on the part of the user [3]–[5]. These methods can provide users with much better segmentation performance than automatic ways.

Fortunately, some specific objects of interest (e.g., face) can be segmented in an unsupervised manner by designing appropriate detectors based on physical model or training scheme [6]. Most work on this topic can be divided into two classes. The first class focuses on the design of robust detector for identifying the specific semantic object, which aims to provide the best candidates for the object of interest. The second class is to present the efficient clustering algorithm for improving the quality of extracting similar pixels. Recently, some joint methods were proposed to perform the semantic object detection and segmentation simultaneously from images [7]. Generally, the final object regions need further modification due to the coarse object segmentation.

Object segmentation based on attention model is another important approach for segmenting semantic object, which allow us to find interesting objects by simulating human visual characteristic. Unlike the previous methods, attention-based scheme aims to segment the meaningful physical entities that are more likely to attract viewers' attention than other objects. Most objects of interest tend to be the attention objects that have distinctive features from their surroundings [1]. A saliency-based visual attention model for rapid scene analysis was first presented in [8], which combined multi-scale image features into a single topographical saliency map. The application of this model on object extraction from color images was reported in [9], which formulated the attention objects as a Markov random field by integrating computational visual attention mechanisms with attention object growing techniques. Based on the visual attention idea, several saliency models are successfully constructed recently to extract the object of interest in different video sequences, such as the facial saliency model [10] and focused saliency model [11].

Although some achievements in semantic object segmentation have been obtained, the limitations of this method are quite evident. It is still a challenge to provide an efficient segmentation solution to extract the semantic object accurately in unsupervised manner, such as the perfect video object planes in MPEG-4 video standard. More work should be carried out to improve the segmentation schemes for meaningful semantic objects.

REFERENCES

[1] H. Li, and King N. Ngan, "Automatic video segmentation and tracking for content-based applications", *IEEE Communications Magazine*, vol. 45, no. 1, pp. 27 - 33, 2007.

[2] T. Meier and K.N. Ngan, "Automatic segmentation of moving objects for video objects plane generation," *IEEE Trans. Circuits and Systems for Video Technology*, vol. 8, no. 5, pp. 525-538, September 1998.

[3] C. Rother, V. Kolmogorov, and A. Blake, "GrabCutl interactive foreground extraction using iterated graph cuts," in *Proc. SIGGRAPH 2004*, pp. 309–314, 2004.

[4] Y. Li, J. Sun, C.-K. Tang, and H.-Y. Shum, "Lazy snapping," in *Proc. SIGGRAPH*, pp. 303–308, 2004.

[5] J. Wang, P. Bhat, R. A. Colburn, M. Agrawala, and M. F. Cohen, "Interactive video cutout," in *Proc. SIGGRAPH*, pp. 585–594, 2005.

[6] H. Li, King N. Ngan, and Qiang Liu, "FaceSeg: automatic face segmentation for real-time video", *IEEE Transactions on Multimedia*, vol.11, no.1, pp.77-88, 2009.

[7] B. Wu, and R. Nevatia, "Simultaneous object detection and segmentation by boosting local shape feature based classifier", *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 17-22 June 2007.

[8] L. Itti, C. Koch, and E. Niebur, "A model of saliency-based visual attention for rapid scene analysis," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 20, pp. 1254-1259, Nov. 1998.

[9] J. Han, King N. Ngan, M. Li, and H.-J. Zhang, "Unsupervised Extraction of Visual Attention Objects in Color Images," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 16, no. 1, pp. 141-145, Jan. 2006.

[10] H. Li, and King N. Ngan, "Saliency model based face segmentation in head-and-shoulder video sequences", *Journal of Visual Communication and Image Representation*, Elsevier Science, vol. 19, no. 5, pp 320-333, 2008.

[11] H. Li, and King N. Ngan, "Unsupervised video segmentation with low depth of field", *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 17, no.12, pp. 1742-1751, 2007.

King N. Ngan received the Ph.D. degree in Electrical Engineering from the Loughborough University in U.K. He is currently a chair professor at the Department of Electronic

Engineering, Chinese University of Hong Kong, and was previously a full professor at the Nanyang Technological University, Singapore, and the University of Western Australia, Australia.

Professor Ngan is an associate editor of the Journal on Visual Communications and Image Representation, U.S.A., as well as an area editor of EURASIP Journal of Signal Processing: Image Communication, and served as an associate editor of IEEE Transactions on Circuits and Systems for Video Technology and Journal of Applied Signal Processing. He chaired a number of prestigious international conferences on video signal processing and communications and served on the advisory and technical committees of numerous professional organizations. He is a general co-chair of the IEEE International Conference on Image Processing (ICIP) to be held in Hong Kong in 2010. He has published extensively including 3 authored books, 5 edited volumes and over 200 refereed technical papers in the areas of image/video coding and communications.

Professor Ngan is a Fellow of IEEE (U.S.A.), IET (U.K.), and IEAust (Australia), and an IEEE Distinguished Lecturer in 2006-2007.

Hongliang Li (M’06) received his Ph.D. degree in Electronics and Information Engineering from Xi'an Jiaotong University, China, in 2005. From 2005 to 2006, he joined the visual signal processing and communication laboratory (VSPC) of the Chinese University of Hong Kong (CUHK) as a Research Associate. From 2006 to 2008, he was a Postdoctoral Fellow at the same laboratory in CUHK. He is currently a Professor in the School of Electronic Engineering, University of Electronic Science and Technology of China. His research interests include image segmentation, object detection and tracking, image and video coding, and multimedia communication system.

Mathematical and Perceptual Models for Image and Video Segmentation

Thrasyvoulos N. Pappas (IEEE Fellow), Northwestern University, USA
pappas@eecs.northwestern.edu

Image segmentation is a key step for a number of problems in image understanding and communications, including content-based retrieval, object-based compression, video summarization, annotation, and surveillance. However, after years of research and considerable progress, it remains a challenging problem. Rather than attempting to cover the vast literature, this paper will present the author's perspective. Our focus will be on the segmentation of natural scenes.

One of the reasons that image segmentation remains such a elusive problem is perhaps that we have set the bar too high. Our expectation is that segmentation algorithms will detect object boundaries. Indeed, the human visual system (HVS) is very good at detecting objects from images of two-dimensional scenes. However, in doing so it relies on a lot of a priori information as, in the majority of cases, only partial information about the object is available to the human retina [1]. A more realistic objective for image segmentation techniques is to detect perceptually uniform regions. As we have argued in [2], such segmentations can be used as medium-level descriptors that can help bridge the gap between low-level features and high-level semantics. Segmentation into perceptually uniform regions can provide semantically significant clues about the object shape and surface characteristics that can lead to object detection and scene analysis.

The basic assumption of a segmentation algorithm is that the image or video sequence consists of regions, each of which has different statistical characteristics, such as color, texture, and shading. However, in natural scenes, the statistical characteristics of perceptually uniform regions are not necessarily uniform, due to variations in illumination, perspective view, and variations in surface properties. In the following, we will examine a hierarchy of algorithms that rely on models of signal characteristics and human perception to overcome such problems. A key feature of these algorithms is that they adapt to local variations in the segment characteristics. To accomplish that, they exploit the fact that the spatial characteristics vary slowly within each segment and rapidly across segment boundaries.

A convenient mathematical formulation for a model-based segmentation algorithm is as a maximization of the a posteriori (MAP) probability density function for the distribution of regions $\boldsymbol{x}$ given the observed image $\boldsymbol{y}$. By Bayes' theorem

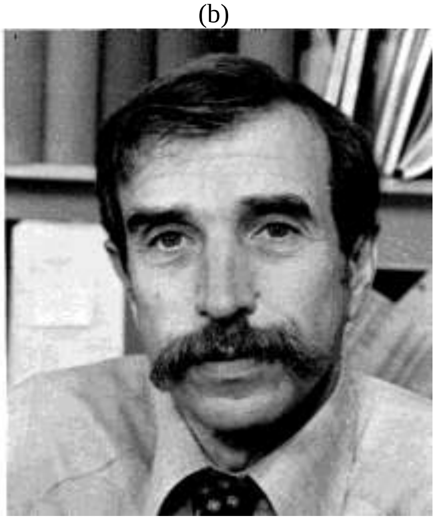
$$p(\boldsymbol{x}|\boldsymbol{y}) \propto p(\mathbf{y}|\mathbf{x})\, p(\boldsymbol{x})$$

where $p(\boldsymbol{x})$ is the a priori density of the region process and $p(\boldsymbol{y}|\boldsymbol{x})$ is the conditional density of the observed image given the distribution of regions. The latter models the texture within each region. For the region process we use a Markov random field (MRF) [3] that models the spatial continuity of the regions by assigning higher probabilities to pixels that have similar labels to their neighbors.

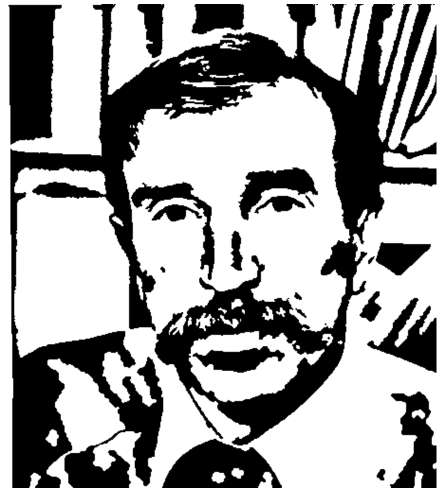
The simplest form of the texture model is to assume no texture, that is, the regions consist of slowly varying intensities plus white Gaussian noise. The adaptive clustering algorithm (ACA) proposed in [3], and extended to color in [4], is an iterative algorithm that generalizes the K-means clustering algorithm [5,6] by adding spatial MRF constraints and adapting to local variations in region intensities. It segments the image into K classes. Each class k is characterized by a spatially varying characteristic function $\mu^k(\boldsymbol{s})$ that replaces the spatially fixed cluster center of the K-means algorithm; here $\boldsymbol{s}$ denotes the pixel location. The algorithm alternates between estimating the characteristic functions and updating the segmentation by maximizing the a posteriori probability density. The key to adapting to the local image characteristics is that ACA estimates the characteristic functions $\mu^k(\boldsymbol{s})$ by averaging over a sliding window whose size progressively decreases; it starts with global estimates and slowly adapts to the local characteristics of each region. A multi-resolution implementation increases computational efficiency and performance. Figure 1 shows examples of ACA with K=4 and K=2.

(a)

(b)



(c)



(d)

Figure 1: Adaptive clustering algorithm: (a) original; (b) 4-level segmentation; (c) original; (d) 2-level segmentation.

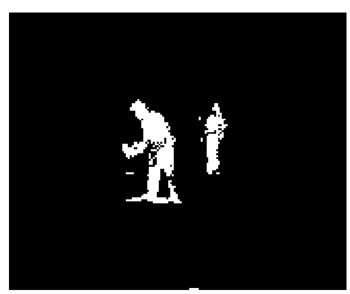
The model we described in the previous paragraph can be generalized to obtain a number of image and video segmentation techniques. First, we consider a spatiotemporal extension of the ACA algorithm. In [7], we used a 3-dimensional MRF to impose spatiotemporal continuity constraints and allowed spatiotemporal variations in the region intensity functions. The temporal constraints and temporal intensity adaptation ensure a smooth transition of the segmentation from frame to frame. However, with the addition of the temporal dimension, the computational complexity increases significantly. A multi-resolution approach and a number of suboptimal implementations can be used to reduce the delay as well as the amount of computation [7]. An alternative approach to dramatically reduce computation without sacrificing performance is to use recursive computation of the local averages [8,9], in a manner analogous to IIR versus FIR filter computations. In fact, we have shown that the recursive approach exploits temporal continuity to decrease the amount of computation per frame relative to ACA applied to each frame independently. The background subtraction algorithms we discuss next can be considered as a special case of spatiotemporal ACA, but it is important to understand their own peculiarities, and in particular, the need for high temporal sensitivity.

The segmentation of a video sequence into foreground (moving) objects and (mostly stationary) background regions is a key task in signal processing and computer vision applications such as video compression and surveillance. The most widely used approach for this task is background modeling and subtraction. This requires building an adaptive model for the background because, in most scenes, its intensity is temporally varying, due to changes in illumination and small movements in the background objects (vibrations, wind, etc.). A number of background subtraction algorithms have been proposed in the literature [10-12]. However, due to computational constraints, the background models are primarily based on temporal statistics and ignore spatial interactions. In contrast, [8,9] proposed a novel spatiotemporal algorithm for joint background subtraction and scene segmentation that exploits spatial dependencies between pixels. As in 3-D ACA, each pixel is modeled by a mixture of spatiotemporal Gaussian distributions, where each distribution represents a different region in the neighborhood of the pixel. The spatiotemporal constraints are controlled by an MRF. By utilizing information in the neighborhood of each pixel, this method obtains accurate and robust models of dynamic backgrounds that are highly effective in detecting foreground objects. Experimental results for indoor and outdoor surveillance videos have shown that it offers significant performance advantages compared to other multimodal methods. An example of joint spatiotemporal segmentation and background subtractions is shown in Figure 2.

Finally, we consider the addition of texture. The adaptive perceptual color-texture segmentation algorithm [13] utilizes two types of spatially adaptive features. The first is based on ACA as a localized description of color composition, and the second uses the energy of the subband coefficients of a steerable filter decomposition with four orientation subbands as a simple but effective characterization of the grayscale component of the spatial texture. Pixels are first classified into smooth and non-smooth classes, and non-smooth pixels are further classified on the basis of dominant orientation, as horizontal, vertical, +45°, -45°, and complex (i.e., no dominant orientation). The segmentation algorithm combines the color composition and spatial texture features to obtain segments of uniform texture. It relies on spatial texture to determine the major structural composition of the image, and combines it with color, first to estimate the major segments, and then to obtain accurate and precise border localization. The border refinement employs an iterative procedure that, like ACA, uses a sliding window of progressively decreasing size to estimate the localized texture characteristics of each region, and compares them to the texture surrounding the current pixel. A spatial MRF constraint is added to ensure region smoothness. Four segmentation examples, superimposed on the original image, are shown in Figure 3.

(a)

(b)



(c)

Figure 2: Joint spatiotemporal segmentation and background subtraction: (a) original; (b) segmentation/background subtraction; (c) foreground.

(a)

(b)

(c)

(d)

Figure 3: Perceptually-tuned color-texture segmentation.

Acknowledgments

This work was supported in part by the National Science Foundation (NSF) under Grant No.CCR-0209006 and the Defense Intelligence Agency (DIA) under DIA/NSF Grant No.IIS-0515929. Any opinions, findings and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of NSF or DIA.

References

1. S. E. Palmer, *Vision Science: Photons to Phenomenology*. MIT Press, 1999.

2. T. N. Pappas, J. Chen, and D. Depalov, “Perceptually based techniques for image segmentation and semantic classification,” *IEEE Commun. Mag.*, vol. 45, pp. 44--51, Jan. 2007.

3. T. N. Pappas, “An adaptive clustering algorithm for image segmentation,” *IEEE Trans. Signal Processing*, vol. SP-40, pp. 901--914, Apr. 1992.

4. M. M. Chang, M. I. Sezan, and A. M. Tekalp, “Adaptive Bayesian segmentation of color images,” *Journal of Electronic Imaging*, vol. 3, pp. 404--414, Oct. 1994.

5. J. T. Tou and R. C. Gonzalez, *Pattern Recognition Principles*. Addison, 1974.

6. Y. Linde, A. Buzo, and R. M. Gray, “An algorithm for vector quantizer

design," *IEEE Trans. Commun.*, vol. COM-28, pp. 84--95, Jan. 1980.

7. R. O. Hinds and T. N. Pappas, "An adaptive clustering algorithm for segmentation of video sequences," in *Proc. ICASSP-95*, vol. 4, (Detroit, MI), pp. 2427--2430, May 1995.

8. S. D. Babacan and T. N. Pappas, "Spatiotemporal algorithm for joint video segmentation and foreground detection," in *Proc. European Signal Proc. Conf. (EUSIPCO*), (Florence, Italy), Sept. 2006.

9. S. D. Babacan and T. N. Pappas, "Spatiotemporal algorithm for background subtraction," in *Proc. ICASSP-07*, vol. I, (Honolulu, Hawaii), pp. 1065--1068, Apr. 2007.

10. K. Toyama, J. Krumm, B. Brumitt, and B. Meyers, "Wallflower: Principles and practice of background maintenance," in *IEEE Conf. Computer Vision (ICCV)*, (Kerkyra, Greece), pp. 255--261, 1999.

11. C. Stauffer and W. E. L. Grimson, "Learning patterns of activity using real-time tracking," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 22, pp. 747--757, Aug. 2000.

12. A. Elgammal, R. Duraiswamy, D. Harwood, and L. Davis, "Background and foreground modeling using nonparametric kernel density estimation for visual surveillance," *IEEE Proceedings*, vol. 90, pp. 1151--1163, July 2002.

13. J. Chen, T. N. Pappas, A. Mojsilovic, and B. E. Rogowitz, "Adaptive perceptual color-texture image segmentation," *IEEE Trans. Image Processing*, vol. 14, pp. 1524--1536, Oct. 2005.

Thrasyvoulos N. Pappas (M'87, SM'95, F'06) (www.eecs.northwestern.edu/~pappas) received the S.B., S.M., and Ph.D. degrees in electrical engineering and computer science from MIT in 1979, 1982, and 1987, respectively. From 1987 until 1999, he was a Member of the Technical Staff at Bell Laboratories, Murray Hill, NJ. In 1999, he joined the Department of Electrical and Computer Engineering (now EECS) at Northwestern University. His research interests are in image and video quality and compression, image and video analysis, perceptual models for multimedia processing, model-based halftoning, tactile and multimodal interfaces.

Dr. Pappas is a Fellow of the IEEE and SPIE. He has served as an elected member of the Board of Governors of the Signal Processing Society of IEEE (2004-2007), chair of the IEEE Image and Multidimensional Signal Processing Technical Committee, associate editor of the IEEE Transactions on Image Processing and the IS&T/SPIE Journal of Electronic Imaging, and technical program co-chair of ICIP-01. He has also served as co-chair of the 2005 SPIE/IS&T Electronic Imaging Symposium. He is currently technical program co-chair for ICIP-09. He has also been selected to serve as Editor-in-Chief of the IEEE Transactions of Image Processing (2010-12). Since 1997 he has been co-chair of the SPIE/IS&T Conference on Human Vision and Electronic Imaging.

Centroid Shifting based Object Tracking

Suk-Ho Lee, Dongseo University, Korea
Moon Gi Kang, Yonsei University, Korea

petrasuk@gmail.com, mkang@yonsei.ac.kr

Object tracking in communication systems

Reliable object tracking algorithms are now being used in a wide variety of applications and communication systems. For example, in video conferences or natural communication with partner robots, human tracking is often used. Furthermore, for the realization of ubiquitous computing environments in everyday life, real-time monitoring and detection of physical events and moving objects over wireless sensor networks are becoming necessary in more and more applications, such as home automation, intelligent surveillance in banks, shopping malls and public buildings, building automation, robotics, vehicular traffic, and navigation. The stability and computational complexity of the algorithm are very important to provide wireless networks and communication systems with reliable sensing data. Various object tracking algorithms have been developed to meet the required conditions. These algorithms vary by the tracking features they use, such as colors, edges, corners, shape, optical flow, texture, etc. The selection of these features depends on several factors, for example, whether or not the video sequence is obtained by a stationary or non-stationary camera.

Object tracking with stationary cameras

If a scene is captured with a stationary camera, then a background reference image can be modeled using statistical background modeling methods [1][2][3][4]. This background reference image can then be subtracted from every incoming frame to identify moving objects. Therefore, motion information such as the optical flow field becomes the dominant feature used in tracking. Whereas the features of the target are used in the estimation of the target position, the dynamics of the target are used to predict the next position of the target and correctly estimate a better one. Kalman filters [5][6], Hidden Markov Model filters [7], and particle filters [8][9] have been applied to predict the target position. Such filters can be applied also to non-stationary cameras if the ego-motion of these cameras is not too large.

Object tracking with non-stationary cameras

The number of applications of object tracking algorithms to low-power mobile systems such as smart phones or small moving machines has increased rapidly. The difficulty is that the tracking for these systems has to be stable, while the complexity of the algorithm has to be kept low due to the limited computation resources in these systems. With non-stationary cameras it is difficult to model background reference images, since the motion information is less reliable than in stationary cameras. If the ego-motion of the camera can be estimated (as in aerial video surveillance applications), then the background can still be modeled by using certain kinds of linear image mapping [10][11]. However, if the camera shows complex motion, then a reference background image cannot be obtained due to the uncertainty of the motion information. In this case, color becomes the dominant feature used in the tracking algorithm. A common approach to this is template matching [12][13]. In this approach, the similarity between the colors in the template, i.e. the target region, in the current frame and the target candidate in the next frame are measured, and the region in the next frame with the highest correlation score is determined to be the next target region. However, this method is very expensive due to the brutal force search required to find the target region in the next frame. Therefore, mean shift-based tracking algorithms have become popular, since they omit brutal searches for the target region by recursively moving to the next target location via a mean shift [14][15][16][17][18][19].

Centroid Shifting Object tracking with non-stationary cameras

Recently, as an alternative to the popular mean shift tracking algorithm, we proposed a new approach where the displacement of the target region is determined not by the mean shift but by the centroid shift of the target colors [20].

The use of centroids has the following advantages:

1) The translation of the target is irrelative to the number of pixels, since the computation of centroids does not depend on the number of pixels.

2) The computational cost is smaller than that in mean shift-based tracking algorithms.

3) The centroids provide for spatiality in color-based tracking and stabilizes object tracking.

4) Object tracking is considered a least square minimization problem with respect to the target location.

An advantage of formulating the tracking problem as a least square minimization problem is that various tracking frameworks can be implemented easily by simply adding appropriate regularizing terms to the energy functional. For example, the Kalman filtering framework can be easily incorporated into the tracking algorithm by simply adding an L_2 -norm difference function of the estimated and the predicted target position to the original distance functional.

Furthermore, since centroids contain information on the size of the target, a size-adaptive algorithm can easily be implemented.

Conclusion

The trade-off between stable performance and low computation cost remains the main problem in object tracking, especially as applications increase. Therefore, we believe that the development of variations and alternatives of mean shift-based tracking algorithms will lead current trends in the R & D of object tracking algorithms for a while. We also believe that centroid shifting-based tracking algorithms will open up a new field of mobile tracking for communication systems due to low computation costs and stability.

Acknowledgements

This work was supported by the IT R&D program of MKE/MCST/IITA. [2008-F-031-01, Development of Computational Photography Technologies for Image and Video Contents.

References

[1] C. Stauffer and W. E. L. Grimson, "Learning patterns of activity using real time Tracking", IEEE Trans. on Pattern Analysis and Machine Intelligence, 2000, 22, (8), pp.747-767.

[2] J. Rittscher, J. Kato, S. Joga, and A. Blake, "A Probabilistic Background Model for Tracking", in European Conf. on Computer Vision, 2000, (2), pp. 336-350.

[3] N. M. Oliver, B. Rosario, and A. Pentland, "A bayesian computer vision system for modeling human interactions", IEEE Trans. on Pattern Analysis and Machine Intelligence, 2000, 22, (8), pp. 831-843.

[4] A. Monnet, A. Mittal, N. Paragios, and V. Ramesh, "Background Modeling and Subtraction of Dynamic Scenes", In IEEE Int. Conf. on Computer Vision, 2003.

[5] S. Julier and J. Uhlmann, "A new extension of the Kalman filter to nonlinear systems," in Proc. SPIE, Vol. 3068, 1997, pp. 182-193.

[6] Y. Boykov and D. Huttenlocher, "Adaptive Bayesian recognition in tracking rigid objects," in Proc. IEEE Conf. on Computer Vision and Pattern Recognition, Hilton Head, SC, 2000, pp. 697-704.

[7] Y. Chen, Y. Rui, and T. Huang, "JPDAF-based HMM for real-time contour tracking," in Proc. IEEE Conf. on Computer Vision and Pattern Recognition, Kauai, Hawaii, Vol. I, 2001, pp. 543-550.

[8] G. Kitagawa, "Non-Gaussian state-space modeling of nonstationary time series," J. of Amer. Stat. Assoc., Vol. 82, pp. 1032-1063, 1987.

[9] R. Van der Merwe, A. Doucet, N. de Freitas, and E. Wan, "The unscented particle filter," Technical Report CUED/F-INFENG/TR 380, Cambridge University Engineering Department, 2000.

[10] D. Murray and A. Basu, "Motion tracking with an active camera", IEEE Transaction on Pattern Analysis and Machine Intelligence, 1994, 16, (5), pp. 449-459.

[11] Q. Yu and G. Medioni, "Map-Enhanced Detection and Tracking from a Moving Platform with Local and Global Data Association", IEEE workshop on motion and video computing, 2007.

[12] R. Marfil, A. Bandera, J. A. Rodriguez, Sandoval, "Real-time template-based tracking of non-rigid objects using bounded irregular pyramids", Proc. of IEEE International Conference on Intelligent Robots and Systems, Vol. 1, pp. 301- 306, 2004.

[13] M. Juza, K. Marik, J. Rojicek, and P. Stluka, "3D Template-Based Single Camera Multiple Object Tracking", Computer Vision Winter Workshop, 2006.

[14] D. Comaniciu, V. Ramesh, and P. Meer, "Kernel-based object tracking", IEEE Trans. on Pattern Analysis and Machine Intelligence', 2003, 25, pp. 564-575.
[15] G. R. Bradski, "Computer vision face tracking for use in a perceptual user interface", Intel Technology Journal, 1998, 2nd quarter.
[16] C. J. Yang, R. Duraiswami, and L. S. Davis, "Efficient Mean-Shift Tracking via a New Similarity Measure", CVPR05, 2005, 1, pp. 176-183.
[17] A. Naeem, S. Mills, and T. Pridmore, "Structured Combination of Particle Filter and Kernel Mean-Shift Tracking", Proc. Int. Conf. Image and Vision Computing, New Zealand, 2006.
[18] K. Deguchi, O. Kawanaka, and T. Okatani, "Object tracking by the mean-shift of regional colour distribution combined with the particle-filter algorithm", Proc. ICPR 2004, pp. 506-509.
[19] G. Hager, M. Dewan, and C. Stewart, "Multiple kernel tracking with Ssd", Proc. IEEE Conf. Comp. Vision Pattern Recognition, Washington, D.C., 2004, 1, pp. 790-797.
[20] S. H. Lee and M. G. Kang, "Motion Tracking Based on Area and Level Set Weighted Centroid Shifting", to appear in IET Computer Vision.

Suk-Ho Lee received his BS, MS, and PhD degrees in electronics engineering from Yonsei University, South Korea, in 1993, 1998, and 2003, respectively. He worked as a researcher in the Impedance Imaging Research Center from 2003 to 2006 and was an assistant professor at Yonsei University from 2006 to 2008. He has been in the Department of Multimedia Engineering at Dongseo University, Busan, South Korea, since 2008, where he is currently a professor. His current research interests include image and video filtering based on partial differential equations, medical imaging and computer vision.

Moon Gi Kang received his BS and MS degrees in electronics engineering from Seoul National University, South Korea, in 1986 and 1988, respectively, and his PhD degree in electrical engineering from Northwestern University in 1994. He was an assistant professor at the University of Minnesota, Duluth, from 1994 to 1997, and since 1997 he has been in the Department of Electronic Engineering, Yonsei University, Seoul, South Korea, where he is currently an associate professor. His current research interests include image and video filtering, restoration, enhancement, and reconstruction. He has published more than 100 technical articles in these areas. Dr. Kang is a member of IEEE and SPIE. He served as the Editorial Board Member for the IEEE Signal Processing Magazine, the Editor of SPIE Milestone Series Volume (CCD and CMOS imagers), and the Guest Editor of the IEEE Signal Processing Magazine Special Issue on Superresolution Image Reconstruction (May, 2003). He has served in the technical program and steering committees of several international conferences. He has also served as the associate editor of EURASIP Journal of Advances in Signal Processing. He is the recipient of 2006 and 2007 Yonsei Outstanding Research Achievement Awards (Technology Transfer), the 2002 HaeDong foundation Best Paper Award and the recipient of the 2000 and 2008 Awards of Teaching Excellence at Yonsei University.

Working with Large Scale Image/Video Databases

B. S. Manjunath (IEEE Fellow), University of California, Santa Barbara, USA
manj@ece.ucsb.edu

Rapid advances in imaging, computing and communication technologies have changed the way people interact with each other. While unimaginable only two decades ago, large personal collections exceeding multiple terabytes of image/video data is common place now. Computing and storage are no longer the issues they were a few years back. More importantly, online collaborative environments, such as flickr, have raised the bar in terms of interactivity and sharing. These are exciting times for multimedia communications, but there are also significant challenges that need to be overcome. In the following, I will outline briefly some of these challenges.

Meaningful Summarization: One of the outstanding challenges in image understanding is to extract and summarize meaningful information in an image/video sequence. Defining what is meaningful is an issue by itself and, in general, is not a well posed problem. Perhaps the best known current work on video summarization is the ongoing TRECVID challenge [1, 2]. The challenges in designing an effective summarization at the semantic level include ways of effectively combining multi-modal features (derived from visual content, camera motion, audio/speech, if available) as they offer complementary information. However, much of the previous work is limited to sampling the video frames and providing the user with a quick snapshot of the larger data rather than creating a semantically meaningful summary. We believe that it is important to first develop robust feature extraction methods for scene and activity analysis before attempting to summarize the video.

Activity Analysis: TRECVID challenge for event detection is a good starting point for getting an overview of the state-of-the-art approaches, and the effectiveness of these current approaches are below acceptable levels for automated methods. TRECVID is somewhat unconstrained in the selection of the data for training and testing, and this could be one of the reasons for the poor performance of these methods. A more recent effort was initiated by DARPA under the VIRAT program (FY09) (see http://www.darpa.mil/ipto/programs/virat/virat.asp for the proposal solicitation) focuses on aerial videos (UAVs) and activity detection in such videos. Some examples of human activities/events that are of interest include walking, running, entering a building or leaving a building, vehicles moving together, etc. While the VIRAT program emphasizes aerial image analysis for defense related applications, the associated image analysis challenges are typical across a broad spectrum of applications.

Scalability, Search and Retrieval: Dealing with large amounts of images/videos also pose significant scalability issues. Besides the sheer magnitude of the raw data that needs to be managed, there is a need for efficient methods that can process the data in a reasonable amount of time and create a rich set of descriptions for search and retrieval applications. Content based image and video retrieval has been an active research topic for almost two decades now, and there are even standardized descriptions such as the ISO/MPEG-7 [4] that one can use to annotate the data. However, these are low level descriptions, such as those directly derived from color, texture and motion information, and the descriptions are typically high dimensional feature vectors ranging in dimensions from 10 to a few hundreds. Searching for similar visual content often requires computing the nearest neighbors in the given feature space. Thus indexing these feature spaces and creating efficient data structures to carry out this nearest neighbor search has attracted much attention in recent years. An associated issue here is the notion of an appropriate distance metric to make the similarity measurements. Even at the level of such low level abstractions of the data, there could be many useful applications such as locating duplicate entries or near duplicate entries in a large database. If one is searching through a public database such as Youtube that may contain tens of millions of video clips, there is a clear need for not only effective descriptions but also efficient indexing schemes to locate such entries in the database. At the Vision Research Lab at UCSB, we have recently developed one such scheme that appears to be quite promising [5], wherein we can search within a database of about 1800 hours of video content local the duplicates in a fraction of a second.

Distributed signal processing: An emerging problem is one in which the data is collected from a large number of distributed and perhaps mobile sensors. This is very typical in many surveillance applications, such as monitoring an airport terminal or a building. Distributed sensing is also very common in environmental monitoring, in emergency response and in creating efficient infrastructures such as smart energy efficient buildings. In each of these cases, the sensors are constrained by available power and bandwidth. For data-heavy video sensors, it is very desirable that the raw data is processed at the sensor and only the "meaningful information" derived from such sensor is then distributed or sent to a central control node. This calls for some kind of coordinated sensing and information processing at the sensor nodes, and this is currently an open research problem. In addition to the camera node computations, other related issues include placement of these cameras to carry out a specific task, rapid deployment of the cameras (e.g., emergency response), tracking over a camera network with incomplete visual coverage, detecting anomalous events in a coordinated manner, and visualizing/summarizing information generated from hundreds of such cameras to aid human analysts. Given that the presence of cameras is a public space is a sensitive issue involving privacy and security, there is also a need for methods that can anonymize the content such that no personal information is collected. These could include, for example, automated schemes that can blur out the faces of the individuals, to more sophisticated methods where the data from such camera networks is no longer the tradition video data but processed information that has a rich visual description but from which it is not possible to create the original video. Finally, there are very few existing testbeds with real data that one can use to develop, test and validate new methods. At UCSB we are building one such testbed with both indoor and outdoor camera networks and we hope to make this data, properly anonymized available to researchers in the near future.

In summary, this is an exciting time for research at the intersection of communications and computing, particularly in the context of image and video analysis. The classical problems of computer vision, such as segmentation, tracking and recognition, still remain. However, with the emergence of large scale, mobile, (and wireless) camera networks and associated applications, there is an immediate need for methods that can process such large amounts of data, extract meaningful summarization/visualization of such data for human analysts, while taking into account the constraints on available bandwidth and computation.

Acknowledgements

Research at the VRL is supported by the following grants: NSF ITR-0331697, NSF III-0808772 , ONR N00014-07-1-0029, ONR N00014-05-1-0816.

References

[1] http://ir.nist.gov/tv2007/tv7rushes/tv7.bbc.devel/. BBC rushes dataset for TRECVID video summarization task - 2007.

[2] P. Over, A. F. Smeaton, and P. Kelly. “The TRECVID 2007 BBC rushes summarization evaluationpilot,” in TVS’07: Proc. of the Intl. Workshop on TRECVID Video Summarization, pages1-15, ACM 2007.

[3] VIRAT program: http://www.darpa.mil/ipto/programs/virat/virat.asp

[4] B. S. Manjunath, P. Salembier, T. Sikora. Introduction to MPEG-7: Multimedia Content Description Language. Wiley International, 2001.

[5] A. Sarkar, P. Ghosh, E. Moxley, and B. S. Manjunath. Video fingerprinting: Features for duplicate and similar video detection and query-based video retrieval. In Proc. of SPIE, Multimedia Content Access: Algorithms and Systems II, volume 6820, 2008.

B. S. Manjunath joined UCSB in 1991 where he is now a professor of Electrical and Computer Engineering. He received his B.E. from Bangalore University in 1985 with distinction, and M.E. (with distinction) in systems science and automation from the Indian Institute of Science in 1987, and the Ph.D. in Electrical Engineering from the University of Southern California in 1991. Professor Manjunath is internationally recognized for his work on image and video databases, texture analysis and segmentation. He and his students have pioneered the development of image texture analysis algorithms and new techniques for image/video segmentation with applications to content based search and indexing of image/video data. He has published over 200 papers in peer reviewed conferences and journals, supervised 20 Ph.D thesis, and he is a co-inventor on 23 US patents. He has co-edited the book on Introduction to MPEG-7 published by Wiley International (2002). He is the Director of the NSF funded graduate education program (IGERT) in Interactive Digital Multimedia and the Director of the Center for Bio-Image Informatics at UCSB. He was an Associate Editor of IEEE Transactions on Image Processing, Trans. Pattern Analysis and Machine Intelligence, Trans. Multimedia and the IEEE Signal Processing Magazine. His research interests include image processing for digital libraries, image analysis, multimedia databases, bio-medical image processing and bio-informatics, data mining and steganography. He is a fellow of the IEEE.

Video Segmentation and Feature Extraction for Communications

Paulo Lobato Correia, Instituto Superior Tecnico, Portugal
plc@lx.it.pt

A Video Communications World

Video communications assume an increasing role in modern lifestyle. Streaming of video contents, stored or live, unidirectional or interactive, is a major trend. This success can be observed in television-like services, by the proliferation of personal communication devices capable of transmitting images and videos and by Internet applications, such as YouTube [1].
A variety of communication channels can be used to access video contents, from wireless channels, supporting simple streaming services or low resolution mobile TV, to direct optical fiber access to a subscribers home, enabling more advanced services, like HDTV or 3DTV.
Despite the differences in the communication infrastructure or in terminal device capabilities, all major video communication applications available still represent the visual information as a sequence of rectangular frames of pixels.
The usage of video analysis techniques, notably segmentation and feature extraction, may allow the offer of improved video communication services, as discussed in the following.

Bringing in Video Analysis

Segmentation is the process of identifying the semantically meaningful components, sometimes called objects, of an image or video sequence.
Unfortunately, a complete theory of video segmentation is not available [2]. Segmentation techniques are often ad hoc in their genesis, differing in the way they compromise one desired property against another. Quoting Pavlidis [3]: "*The problem is basically one of psychophysical perception, and therefore not susceptible to a purely analytical solution. Any mathematical algorithms must be supplemented by heuristics, usually involving semantics about the class of pictures under consideration*". Additionally, the application context has an important role for supplying useful heuristics. For instance, it is not easy to segment static, low contrast, objects. On the other hand, moving objects may be easier to detect.
Segmentation of objects in images and video sequences can be a crucial task for a wide variety of multimedia applications, such as object-based video coding, manipulation and identification.

The semantic information provided by a segmentation algorithm can be complemented by feature extraction, to support advanced communication services.
A feature can be understood as a distinctive characteristic of data that signifies something to somebody. Some features are related to an higher level of abstraction than others; their extraction may rely on lower-level features, or may eventually require user interaction. One such example is the estimation of an object priority, which typically relies on lower-level spatial and temporal features, such as size, shape, orientation, position in the scene, motion activity and stability in time, as well as on high level features such as the object type. Also the title of a movie or the name of an actor are examples of features.
Yet another level of understanding can be attained by affective video content analysis, which tries to account for the type and amount of feelings or emotions associated to a given portion of video. This goes one step beyond the cognitive level achieved with segmentation and feature extraction. Affect can be modelled by 3 components: valence (a positive or negative experience), arousal (the intensity of an emotion) and dominance (to distinguish emotional states with similar valence and arousal values) [4].

Analysis outputs can be used to adjust the video contents to be transmitted to particular users, according to their preferences, communication channel or terminal device capabilities in use.
This can be achieved, for instance, through content adaptation, using analysis to create a video summary that fits the user request and usage environment characteristics [5].

Using Video Analysis in the Encoder

Given the availability of object-based video coding and description standards, such as MPEG-4 [6] and MPEG-7 [7], the development of applications exploiting them critically depends on the availability of segmentation algorithms able to identify relevant objects.
In some controlled environments automatic segmentation can be easily achieved, e.g., using chroma-keying techniques. In other cases, the segmentation is inherently present, as when dealing with synthetic contents.

But, also for rectangular video applications the availability of segmentation and feature extraction information about the contents can play a role that makes a difference.

In communications environments, the video analysis information can be used to control encoder parameters. Encoders are non-normative, as standards only need to specify the syntax and semantics of the bitstream, together with the decoding process. Encoders are, therefore, open to technical evolution and competition among different manufacturers – often the encoding details are not fully revealed. Modules like rate control, error resilience or pre-processing can benefit from the available segmentation and feature extraction information.
The control of the encoder should be as much as possible automatic, benefiting from the video analysis system outputs.

Traditional frame-based encoders may devote a larger amount of coding resources to image areas considered more important. For instance, in video telephony, the segmentation of the talking person's head and shoulders can be explored to ensure a better image quality in those areas, when compared to the background [8].
Encoders may adjust the spatial detail, e.g., by varying the macroblocks' quantizer step size or the number of encoded DCT coefficients. The adequate spatial detail can be determined by features like homogeneity, flatness or the presence and type of edges. Also, areas of interest within images can be identified according to their relevance [9].
Also the temporal resolution, i.e., the frame rate, can be adjusted, by skipping some of the input frames, based on content analysis results such as the detected motion activity.

In object-based video, objects may be separately encoded, making more obvious the advantage of using information about the objects' features to control the encoder and, ultimately, to achieve a better subjective video quality. The number of objects to transmit can be controlled, as well as the content's semantic resolution, understood as the number of independently transmitted objects, e.g., deciding to merge the less important objects. Given the set of objects to encode, the available resources must be distributed among them and within each of them. In particular, the precision with which the shape and texture data are encoded must be managed.

Besides controlling the distribution of coding resources among image areas (or objects), segmentation and feature extraction information can also be used to make the video stream more adequate for the usage environment [10]. For instance, the amount of error protection information associated to each object or image area can be adjusted, e.g., by controlling the frequency of resynchronization markers.
Also the ability to select appropriate images, or object images, to use as references for temporal prediction (e.g., I- or P-frames in MPEG terminology), can benefit from the analysis work.

When the applications require scalable encoding, decisions on how many and which scalability layers are appropriate for each object, or for the complete frame, as well as the type of scalability to select (e.g., spatial or temporal), are needed.
With the advent of the SVC standard [11], scalability is assuming a major role, avoiding the need for servers to host multiple versions of the same content, encoded with different bit-rates, e.g. using transcoding techniques.
An example of this trend includes the generation of scalable bitstreams organized according to storyboards or video skim strategies, allowing to adjust the playback duration [12]. In another example, video content analysis detects events of interests in a video; the more relevant temporal segments can then be encoded with higher spatio-temporal quality [13].

In summary, there is a lot of to be gained by the usage of video analysis tools, notably to personalize video coding according to the operation environment.

References

[1] http://www.youtube.com/
[2] R. Haralick, L. Shapiro, "Computer and robot vision", Vol. I, Addison-Wesley, 1992
[3] T. Pavlidis, "Structural pattern recognition", Springer-Verlag, 1977
[4] A. Hanjalic, L.-Q. Xu, "Affective video content representation and modeling", IEEE Trans. on Multimedia, Vol. 7, N. 1, pp. 143-154, Feb. 2005
[5] P. Fonseca, F. Pereira, "Automatic video summarization based on MPEG-7 descriptions", Signal Processing: Image Communication, Vol. 19, N. 8, pp. 685-699, Sept 2004
[6] ISO/IEC 14496, "Information technology - Coding of audio-visual objects", 1999
[7] ISO/IEC 15938, "Multimedia content description interface", 2001

[8] H. Luo, A. Eleftheriadis, J. Kouloheris, “Statistical model-based video segmentation and its application to very low bit-rate video coding”, Signal Processing: Image Communication, Vol. 16, pp 333-352
[9] P. Correia, F. Pereira, “Video object relevance metrics for overall segmentation quality evaluation”, EURASIP Journal on Applied Signal Processing, May 2006
[10] L.D. Soares, F. Pereira, “Adaptive shape and texture intra refreshment schemes for improved error resilience in object-based video coding”, IEEE Trans. on Image Processing, Vol. 13, No. 5, pp. 662 - 676, May 2004
[11] H. Schwarz, D. Marpe, T. Wiegand, “Overview of the scalable video coding extension of the H.264/AVC standard”, IEEE Trans. on Circuits and Systems for Video Technology, Vol. 17, N. 9, pp. 1103-1120, 2007
[12] L. Herranz, J. Martínez, “An integrated approach to summarization and adaptation using H.264/MPEG-4 SVC”, Signal Processing: Image Communication, Vol. 24, pp. 499–509, 2009
[13] N. Ramzan, T. Zgaljic, E. Izquierdo, “An efficient optimization scheme for scalable surveillance centric video communications”, Signal Processing: Image Communication, Vol. 24, pp. 510–523, 2009

Paulo Lobato Correia graduated as an Engineer and obtained the MSc and PhD degrees in Electrical and Computer Engineering from Instituto Superior Técnico (IST), Universidade Técnica de Lisboa, Portugal, in 1989, 1993, and 2002, respectively. He is currently Assistant Professor at the Electrical and Computer Engineering Department of IST. Since 1994 he is a researcher at the Image Group of Instituto de Telecomunicações. He participated in several national and international research projects, currently coordinating project BIONSE (Biometrics with Enhanced Security using Distributed Source Coding).

In 2004 he was General Co-Chairman of the International Workshop on Image Analysis for Multimedia Interactive Services (WIAMIS’2004); in 2007 he was Program Chair for the Picture Coding Symposium (PCS’2007). Until Dec. 2008 he was an elected member of the EURASIP (European Association for Signal Processing) Administrative Committee.

He is an Associate Editor for "IEEE Transactions on Circuits and Systems for Video Technology" and a member of the EURASIP/Elsevier Signal Processing Journal editorial board. He is the Portuguese representative in the Information and Communication Technologies (ICT) Domain Committee of the COST (European Cooperation in Science and Technology) research program.

Current research interests include video analysis and processing, namely video segmentation, objective video segmentation quality evaluation, content-based video description, video communications, biometric recognition, underwater image analysis and road pavement surface image analysis.

Rate Control for Efficient Video Communications

Paulo Nunes, Instituto Superior de Ciências do Trabalho e da Empresa, Portugal
Fernando Pereira (IEEE Fellow), Instituto Superior Técnico, Portugal
paulo.nunes@lx.it.pt, fp@lx.it.pt

The Problem

Since most precious resources in life are typically scarce, it is essential to find some optimal or at least close to the optimal way to use the resources available following in practice the proverb that "A penny saved is a penny gained". This target implies allocating the available resources in order the relevant objectives are reached in the best way, for example by maximizing one or more performance criteria. As stated in Shannon's seminal paper [1], "*The fundamental problem of communication is that of reproducing at one point either exactly or approximately a message selected at another point*". Considering that communication (transmission or storage) resources are scarce, Shannon's problem can be split into two main parts: What information should be transmitted? How should it be transmitted? These questions may easily be transposed for today's image, video, and audio encoders which have precisely the task of deciding which coding symbols should be used and how they should be translated into bits to represent in the most efficient way a certain content under some relevant constraints, for example the available rate or a target quality. In reaching this target, content analysis plays an important role since the more it is known about the content, the better may be the encoder decisions in terms of rate allocation.

The Solution: Rate Control

The tradeoff between the compression ratio (or equivalently the used rate) and the distortion is inherent to any lossy compression scheme, i.e., a low distortion requires a high rate (low compression ratio) and vice-versa. Rate-distortion (RD) optimization, which is typically seen as the selection of the set of encoding parameters minimizing some cost function that expresses a rate-distortion trade-off, is far from a trivial task. In fact, complexity is a key factor in RD optimization techniques. Two major sources of complexity in this process are: i) the collection of relevant RD data, which may involve several encoding/decoding operations; and ii) the selection of the best RD operation point for the conditions and constraints at hand. The methods to define the allocation of rate resources (i.e., the bits) to the information to be coded meeting the requirements of a transmission channel (e.g., bit rate), a storage device (e.g., throughput, capacity), or an application (e.g., minimum quality or delay) are well known as rate control or bit allocation methods. In practice, these methods imply analyzing the data to be coded in order the available resources are used in a knowledgeable way; the better, and typically more complex, the analysis, the more efficient is the rate control process.

Rate Control in Video Coding Standards

While rate control is a rather old problem in the information theory community [2], its importance has significantly grown with the advent of the multimedia age, starting in the early nineties of the last century. In fact, in the image, video, and audio coding domains, adequate rate control methods are central in determining the quality achieved in the context of a certain coding scheme. For the explosion of the multimedia age, coding standards, notably those created by JPEG, MPEG and ITU-T, played a major role. Interestingly, rate control methods although strongly determining the quality or distortion achieved with those standards for a certain rate or for a certain quality are not normatively specified; this means they are non-normative and thus the rate control solution to be used is free in the context of those standards. The reasoning is simple: since a standard is always a constraint of freedom, it is important to make them as minimally constraining as possible while still providing interoperability in order the standard may evolve as much as possible through the non-normative areas. These non-normative areas correspond to the tools for which normative specification is not essential for interoperability, which is the case of bit rate control. Whatever the rate control solution is used at the encoder, the decoder will always be able to decode the received bitstream provided it is syntactically and semantically compliant; naturally, the reached quality/distortion will not be the same for all rate control methods and thus the interest in keeping developing more efficient rate control solutions. On the contrary, the normative tools are those specified by the standard and whose specification is essential for interoperability. The consequence is that better non-normative tools can always be used, even after the standard has been finalized, and thus it is possible to rely on industry competition for continuously obtaining better results. In fact, it will be through the non-normative tools that products will distinguish themselves, which only reinforces the importance of this type of tools and thus of rate control technologies. While rate control solutions are not determined by the standards, making them even more important in terms of distinguishing interoperable products in the marketplace, the application domain and conditions in question may put significant constraints on the rate control solution to adopt. For example, it is clear that the lower is the available rate, the more critical is the rate control performance since the scarcer are the resources the better they have to be used. It is also clear that real-time and off-line bit rate controls are rather different solutions, answering to rather different rate allocation problems. Generally, the more information is available about the data to encode (typically much more in off-line coding), the more efficient may be the rate allocation, naturally at the cost of some extra complexity, and eventually delay, to collect and process the relevant RD data.

The Rate Control Challenges

The fact that video encoders may accept nowadays a growing amount of complexity motivates the increasing usage of optimization techniques such as the Lagrangian optimization adopted in the H.264/AVC (Advanced Video Coding) video coding standard reference software. However, research on rate control is far from being finished. Besides the continuous search for better RD modeling for the available and emerging coding schemes or more efficient optimization techniques, rate control did not yet considered enough the impact of the human visual system in their modeling and optimization. For example, performing exhaustive optimization for a mean square error distortion metric may reveal itself rather meaningless for specific types of images and video where this metric is less good in terms of expressing a subjective quality evaluation; and the subjective impact is always the last assessment criteria giving expression to the proverb stating that the “truth is in the eye of the beholder”.

A major trend in video coding is scalable coding, layered or not, like in the Scalable Video Coding (SVC) standard and multiple description coding (MDC). Scalable video coding is very adequate to deal with heterogeneous environments, notably different transmission channels and terminals. This focus is justified by the growing variety of available networks, both wired and wireless, with many different features, and with a large range of connection bandwidths. For some of them, the available bandwidth varies in a very dynamic way, and transmission errors, congestions and information losses can occur in a very unpredictable way. Also the capabilities of the user devices may have an important role, since they have very different characteristics, notably in terms of spatio-temporal-quality resolution, memory, computational power, and screen size. While for layered scalable video coding it may be essential to develop rate control solutions able to optimize the distortion for each layer considering a certain rate and vice-versa, for multiple description coding efficient bit allocation techniques are essential to define which data, e.g., transform coefficients, go within each description and with which level of distortion in order to minimize the redundancy between descriptions while maximizing the global and single description qualities. Additionally, since error-prone channels may cause corruption or loss of data, when some feedback about the network conditions is available at the encoder (e.g., average instantaneous bit rate and packet loss rate), this knowledge can be used to drive the encoder allowing coding efficiency to be effectively combined with error robustness by controlling the amount of Intra coding refresh in the bitstreams; this may mitigate the effects of errors, reducing error propagation.

After AVC and SVC, the emerging Multiview Video Coding (MVC) standard, which efficiently codes the multiple video views for a visual scene, is bringing another dimension to rate control since the rate resources must also be efficiently distributed among the various views depending on the amount of temporal and interview redundancy; this rate allocation is closely connected to the type of prediction structures used in MVC encoding depending on the compression efficiency and random access constraints.

In the context of object-based video coding, which still waits to explode in terms of successful applications, rate control still has a long way to go to fully consider the specific characteristics of each object in the scene and optimize the shape versus texture distortion if departure from the most typical lossless shape model is to be accepted. When dealing with objects, semantics may play a significant role and thus impact on the rate allocation since objects with different semantic value very likely deserve/require different distortions. The semantic dimension of the object-based representation paradigm opens new degrees of freedom to rate control, yet to be explored, such as the semantic resolution control and the amount of content control. The semantic resolution of a certain scene is related to the semantic detail provided; this means to the number of objects in which a certain amount of video data is organized (for the same amount of pixels, more objects mean higher semantic resolution and vice-versa). The amount of content in a scene is related to the number of objects and to the corresponding amount of data; decreasing the amount of content means reducing the number of objects in the sense that the pixels corresponding to a certain object are removed (not in the sense that objects are merged).

Finally, it must be recognized that, although the rate control problem can most of the times be simply seen as a rate-distortion optimization problem, the theoretical rate-distortion bounds depend on arbitrarily long data blocks and, eventually, unlimited computational power. Therefore, considering as the single limited resource the rate associated to the transmission channel or storage device may be a rather simplistic approximation. In a world where limited capacity devices such as mobile terminals have a rising importance, other resources such as the computational resources, memory, and the battery life have also to be recognized as scarce. This type of recognition implies that the relevant problem to solve is not anymore only an RD optimization problem but rather an RDC (complexity) problem to better express the real world limitations.

In conclusion, video coding rate control is continuously opening new frontiers and dimensions, determining efficiency for video communications.

Reference

[1] C. E. Shannon, “A Mathematical Theory of Communication”, Bell System Technical Journal, vol. 27, pp. 397-423, 623-656, 1948.

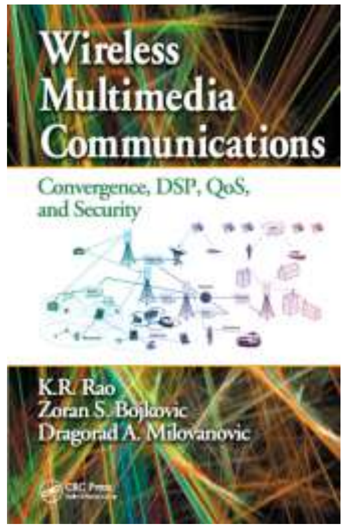
[2] C. E. Shannon, “Coding Theorems for a Discrete Source with a Fidelity Criterion”, IRE National Convention Record, vol. 7, part 4, pp. 142-163, NY, March 1959.

Paulo Nunes (M’07) received the B.S., M.Sc. and Ph.D. degrees in electrical and computer engineering from the Instituto Superior Técnico (IST), Universidade Técnica de Lisboa, Portugal, in 1992, 1996, and 2007, respectively. He is currently a Professor at the Information Science and Technology Department, Instituto Superior de Ciências do Trabalho e da Empresa (ISCTE), Portugal, and a member of the Research Staff of the Telecommunications Institute, Portugal. His current research interests include object-based and frame-based rate control and video coding. Dr. Nunes has participated in various European projects and has acted as project evaluator for the European Commission. He acts often as reviewer for various conferences and journals and has contributed more than 30 papers.

Fernando Pereira is currently Professor at the Electrical and Computers Engineering Department of Instituto Superior Técnico. He is an Area Editor of the Signal Processing: Image Communication Journal and is or has been an Associate Editor of IEEE Transactions of Circuits and Systems for Video Technology, IEEE Transactions on Image Processing, IEEE Transactions on Multimedia, and IEEE Signal Processing Magazine. He was an IEEE Distinguished Lecturer in 2005 and IEEE Fellow in 2008. He has been a member of the Scientific and Program Committees of many international conferences and has contributed more than 200 papers. He has been participating in the work of ISO/MPEG for many years, notably as the head of the Portuguese delegation, Chairman of the MPEG Requirements Group, and chairing many Ad Hoc Groups related to the MPEG-4 and MPEG-7 standards. His areas of interest are video analysis, processing, coding and description, and interactive multimedia services.

Editor's Selected Book Recommendation

K. R. Rao, Z. S. Bojkovic, D. A. Milovanovic. *Wireless Multimedia Communications: Convergence, DSP, QoS, and Security*, CRC, Nov. 2008.



Overview

Wireless multimedia communication is one of the major themes in today's information communication technologies, which enriches human beings' life greatly. Since it contains various aspects from wireless networking to multimedia communication, signal processing and security, it attracts various practicing engineering or research working in the past decade.

This book, authored by experienced experts, reviews the recent development in wireless multimedia communication, provides the underlying theory, concepts and principles related to the field, and presents some latest works. The covered topics include the evolution of wireless networks, various convergence technologies, wireless video techniques, wireless multimedia services and applications, wireless networking standards, cross-layer wireless multimedia techniques, mobile Internet, and evolution toward 4G networks. Its wide coverage makes it suitable for a teaching textbook, research reference or engineering handbook.

Book Content

The book is composed of 9 chapters. Chapter 1 gives a brief introduction to wireless networking, including the evolution of mobile networks, history of third generation systems, evolving wireless multimedia networks, and typical wireless multimedia services.

Chapter 2 provides an overview of the key convergence technologies to offer many services from the network infrastructure point of view. It is composed of a short presentation of the next generation network architecture, convergence technologies for third generation (3G) networks, technologies for 3G cellular wireless communication systems, the 3G wideband code-division multiplex access (WCDMA) standard, packet data broadcast services, and challenges in the migration to fourth generation (4G) mobile systems.

Chapter 3 surveys the wireless video that has been commercialized recently. It consists of a general framework of energy-efficient wireless video communication system, rate control in wireless streaming video, content delivery technologies in difference layers, and the H.264/AVC standard in the wireless video environment.

Chapter 4 shows the current issues and challenges in the field of wireless multimedia services and applications. The considered services include real-time IP multimedia services, Internet Protocol multimedia subsystem, and location-based services.

Chapter 5 summarizes specifications for wireless networking standards that support a broad range of applications, e.g., wireless local area networks (WLANs), wireless personal area networks (WPANs), wireless metropolitan area networks (WMANs), and wireless wide area networks (WWANs). The reviewed standards include IEEE 802.X standards, WLAN link layer standards, Wireless asynchronous transfer mode LAN, European Telecommunication Standard Institute (ETSI) BRAN HIPERLAN standard, etc. Additionally, H.264 video transmission over IEEE 802.11 is introduced.

Chapter 6 focuses on advances in wireless video. It introduces the error robustness support using the H.264/AVC standard, provides an overview of error resilience video transcoding for wireless communications, highlights recent advances in joint source coding and optimal energy allocation, and analyzes the multipath transport of video streams.

Chapter 7 concentrates on cross-layer wireless multimedia design. It describes the cross-layer architecture for video delivery over wireless channel, cross-layer optimization strategy, a short overview of cross-layer design approaches for resource allocation in 3G CDMA network, and the cross-layer resource allocation for integrated voice/data traffic in wireless cellular networks.

Chapter 8 focuses on mobile Internet and related technologies. It presents and analyzes the protocols for mobile Internet, describes IP mobility for cellular and heterogeneous mobile networks, introduces the scalable application-layer mobility protocols, reviews the mobility and QoS, and presents the network architecture for seamless mobility services.

Chapter 9 discusses the evaluation of toward 4G networks. The covered topics include migration to 4G mobile systems, beyond 3G and toward 4G networks, heterogeneous 4G system integration and services, all-IP 4G network architecture, and QoS and security for 4G networks.

Conclusion

The book's chapters have different properties targeted for various readers. For example, Chapters 3, 6, 7 and 9 can be used for textbook teaching material, Chapters 3, 6 and 7 keep pace with rapid development of multimedia communication, and contains valuable information for research, and Chapters 1, 2 ,3, 8, and 9 provide rich information to engineers. Overall, it keeps pace by including the latest developments, and will provide various readers with a valuable tool and resource in wireless multimedia communications. I would recommend this book to graduate students who want to familiarize themselves with various topics in wireless multimedia communication, and to experienced researchers and engineers as a reference book.

** The Column Editor recommending this book is Dr. Shiguo Lian.*

Focused Technology Advances Series

Directional Link Networks: Enabling Technologies for Multimedia Sensor Networks

Deepa Kundur, Texas A&M University, USA
deepa@ece.tamu.edu

The vision of ambient intelligence consists of a multitude of electronic devices and sensors that are seamlessly embedded into people's daily life. Currently, the most promising applications for this environment include home entertainment, healthcare, monitoring, automation, while it is Classically, wireless sensor networks have been envisioned to consist of groups of lightweight sensor nodes that observe *scalar* data, communicate wirelessly, and are densely distributed, collaborative, autonomous, hierarchical and secure. The nodes are distributed in a physical region containing a phenomenon of interest, which is to be monitored and possibly controlled. When the sensor nodes collect diverse types of information such as temperature, humidity, acoustic and visual data simultaneously, they are termed "multimodal sensors". Multiple types of sensing can occur within the same node through the use of distinct sensing technologies or across different nodes each having a single, but distinct sensor type. Multimodal sensors that collect multimedia information such as digital images, video and audio form a multimedia sensor network (MMSN).

Multimedia Sensor Networks: MMSNs represent a form of wireless sensor network in which a subset of sensors often collect higher bandwidth content; MMSNs that sense and process visual information will, in particular, play a critical role in the world's advancement, security and well-being. They can help interface to existing video surveillance infrastructure. For applications including healthcare surveillance, environmental observation and vehicle control, visual and other forms of broadband data are crucial for monitoring. MMSNs can be used for critical tasks often performed by humans such as the monitoring of sick patients. The rich visual signatures of surface currents for oceanographic monitoring makes MMSNs a cheaper alternative to characterize full ocean water columns. In situations where the network *sink* is a human observer, processed visual data from the network can enhance user-interactivity; for example, for unmanned ground or aerial vehicles MMSNs provide the feedback necessary for human operators for make critical motion and target decisions. The proliferation of low-cost portable off-the-shelf media sensing devices has motivated the recent development of vision-rich MMSN system theory, architectures and test beds.

MMSNs possess unique design challenges. First, in contrast to scalar networks, MMSNs require high speed hierarchical networking capabilities to transport broadband data; the improved scalability provided by employing a more hierarchical and power-specialized node architecture is especially advantageous when higher bandwidth communications is involved. Second, MMSNs are heterogeneous where nodes fall in classes with distinct sensing capabilities; for example, scalar sensors such as motion detectors can trigger vision acquisition and the associated traffic patterns may be bursty. Third, given the safety-critical applications facilitated by MMSNs, security and privacy within such networks are of significant concern.

Directional Link Networks: Directional link networks have recently shown potential to address the unique challenges of MMSN systems. Employing directional links provides advantages over traditional omnidirectional transmission for ad hoc sensor networks. By focusing energy in one direction, the potential for spatial reuse is increased while the consumed power and interference are reduced for the same transmission radius; this lengthens network lifetime while providing increased signal strength and reduced multipath components. Similarly, for the same power consumption, longer communicate ranges or higher bandwidth can be achieved facilitating multimedia communications and bursty traffic patterns. Furthermore, security is enhanced due to the reduced spatial signature of the communication signal from a broadcast disk-based model (for omnidirectional communications) to a sector-inspired model, thereby reducing the chances of eavesdropping potentially providing inherent security and privacy. Given these physical layer advantages, there is currently research interest in evaluating directional link technologies for advanced high speed networking systems.

Directional Links at the Physical Layer: Two main technologies exist for directional link communications: free space optical (FSO) and directional radio frequency (RF). **Figure 1** illustrates the idealized differences among the physical layer communication footprints of traditional omnidirectional RF, directional RF and FSO approaches.

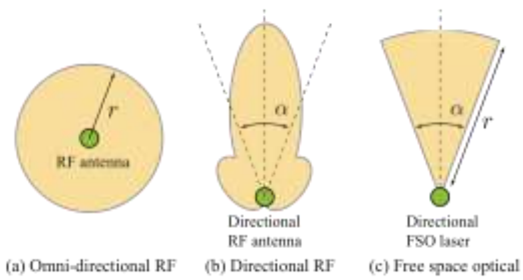


Figure 1. Communication footprints for omnidirectional, directional RF and FSO transmission

The traditional omnidirectional RF paradigm is currently employed in most wireless networking applications; it has the advantage of simplicity for networking protocols (since direction of transmission does not have to be effectively synchronized with other nodes) and improved connectivity given the broadcast nature of communications. However, the bandwidth-power consumption tradeoff is not competitive for MMSNs necessitating wired communication solutions where possible. However, in many applications for which wireless communications is a necessity (e.g., ad hoc networking in geographically remote regions), the link speeds offered by directional RF and FSO technologies demonstrate great potential thus warranting further study.

In the case of FSO communications, the potential for highly compact size (dust-like as proposed for the original Smart Dust) and power efficiency in comparison to RF communications makes them highly favorable for MMSNs. However, atmospheric conditions such as fog, clouds, snow and rain affect link reliability that must be addressed through physical layer processing and network robustness. Furthermore, the line of sight nature of communications makes transceiver alignment a significant issue especially in ad hoc networking contexts where communication may be impeded by physical objects such as buildings or walls. RF communications, in contrast, does not suffer from line-of-sight (LoS) constraints. However, the need for multiple antennas for transmission and/or reception results in node that may be impractically large or costly for lightweight MMSNs. Moreover, there is a beam steering delay that must be accounted for during networking.

Much existing research on directional links has focused on physical layer considerations to maintain bandwidth and security. However, as these devices are connected, networking challenges must also be addressed. For example, coding, modulation and signal processing strategies for various transceiver configurations can significantly improve link quality. However, the existence of the directionality of links due to *node deafness* (i.e., a node s_a cannot be heard by a node s_b within its proximity because s_b's receiver is directional) or *node invisibility* (i.e., s_a cannot transmit s_b because s_a's transmitter is directional) raises fundamental design questions at the networking level. To exploit the physical layer advantages of directional communications network layer mechanisms must be carefully designed to account for a "multi-hop view". In such a context medium access control and routing performance may not improve proportionally to the link speeds due to overhead.

Medium access control strategies must account for any steering involved during transmission and/or reception for directional RF nodes when communicating with immediate neighbors. Temporary node deafness and invisibility results in overhead due to the need for network reconfiguration. On a larger scale, network connectivity and routing issues must be considered. The traditional challenges of reliability, throughput and security must be studied in this new context. Not only does analysis of directional link networks provide performance bounds for emerging MMSNs, but in standard heterogeneous networks in which different devices have distinct communication ranges, directional links may be accommodated to avoid under-utilization and to diminish standard overhead costs.

Connectivity for Directional Networking: The range extension of directional communications can improve the LoS connection between two geographically distant nodes. However, questions arise as to the implications of directional links to *network connectivity*. Network connectivity for standard bidirectional-link networks requires that at least one sequence of nodes (i.e., a path) exist connecting every possible node pair. For directional networks a notion of a *strongly connected* network is needed. Specifically, a network is strongly connected if for every node pair (s_a, s_b), paths from s_a to s_b and from s_b to s_a exist. For example, **Figure 2** illustrates a unidirectional network that is strongly connected. In contrast to bidirectional links, one sees that the paths from s_a to s_b and from s_b to s_a are necessarily distinct.

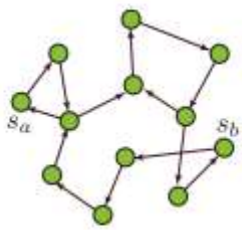


Figure 2. Strongly connected directional link network

The connectivity of a MMSN employing directional links is dependent on the transceiver configuration. Ideally, four transceiver configurations are possible for each directional RF or FSO node as detailed in Table 1. For the omni-omni case, it is clear that there is ideally no issue with unidirectional links. In all other configurations node deafness and/or invisibility is possible causing unidirectional links.

Table 1. Wiress transceiver configurations.

transmitter	**receiver**	**Node deafness/ invisbility**
omni	omni	neither
directional	omni	node invisibility
omni	directional	node deafness
directional	directional	both

In the remainder of this article, we will focus on the directional-omni transceiver case that is a common model for sensor networks such as Smart Dust that uses FSO communications. Here the nodes are static and are assumed to be randomly deployed in a 1-km by 1-km square geographical region with random position and orientation. Thus, a node can receive information via its omnidirectional receiver if it is within the LoS (i.e., static beam) of another node. **Figure 3** shows a possible realization of such a model for 200 nodes. Given the random nature of the associated network graph, probabilistic methods are used to assess connectivity. Three parameters, the number of nework nodes *n*, communication range *r* and beam width α, characterize the properties of the associated *random graph*. **Figure 1**(c) illustrates *r* and α in the context of the transmission sector of an FSO node.

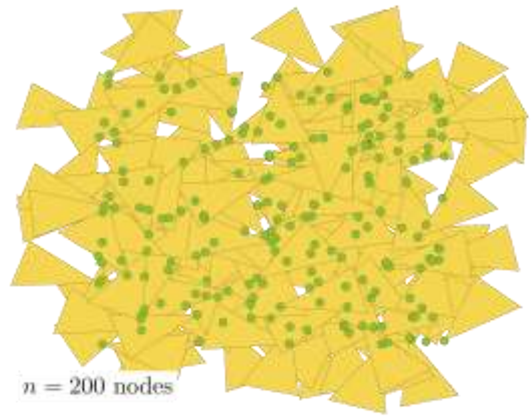


Figure 3. Randomly deployed unidirectional link network

Connectivity is therefore analyzed in terms of what is classically termed the *parameter assignment problem*. The specific problem is to determine the parameters (n, r, α) that guarantee at least a certain probability of connectivity of the associated random graph. Finding an exact expression for this probability of connectivity as a function of (n, r, α) is an open problem. Thus, research bounds this likelihood from above with the probability that there is *no isolated node*.

Node Isolation vs. Network Connectivity: A network node is isolated if it cannot transmit to or receiver from another node in the network. The situation when no network node is isolated not equivalent to the case of a directional network being strongly connected. For example, **Figure 4** shows a situation in which there are no isolated nodes (i.e., every node can communicate to at least one node and is able to receive from at least one node). However, the directional link between s_b and s_a that connects the two loops as well as the partition imply that the network is not strongly connected.

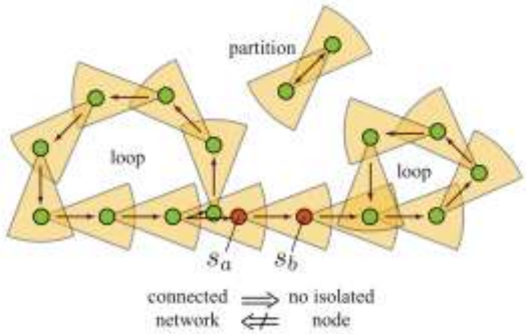


Figure 4. Example of a network with no isolated node that is not strongly connected

An analytic expression for the probability of no isolated node (representing an upper bound on the probability of connectivity) has been derived by the author to be:

$$p_d = \left[1 - e^{-\frac{n\alpha r^2}{2}}\right]^n \left[1 - \frac{e^{-\frac{n\alpha r^2}{2}}}{1 - e^{-\frac{n\alpha r^2}{2}}}\left(1 - \frac{\alpha r^2}{2}\right) \cdot \left(e^{\frac{n\alpha r^2(2\pi-\alpha)}{2\pi(2-\alpha r^2)}} - 1\right)\right]^n$$

Figure 5 compares the probability of connectivity and no isolated node for 500 nodes. The solid pink line is the probability of connectivity empirically obtained through the averaging of 1000 random deployment realizations and associated test for connectivity. The solid blue line is the analytic expression for the probability of no isolated node shown above, which as predicted is an upper bound. The red line represents the simulated probability of no isolated node and the black dash-dot line is the probability of connectivity compensating for edge effects using a Toroidal distance measure instead of Eucledian. Similar results are found for larger values of beam width α as illustrated in **Figure 6** and **Figure 7**. The latter graph corresponds to the bidirectional communication model that is commonly employed for omnidirectional RF wireless ad hoc networks. In all cases, the probability of no isolated node represents an upper bound on the probability of connectivity. As the beam width grows, this bound naturally tightens. Furthermore, one sees that compensating for edge effects also diminishes any differences between the probabilities. It should be mentioned that hierarchy where a randomly selected subset of nodes (e.g., cluster heads) are connected bidirectionally to one another can be shown to significantly improve connectivity; however, this concept is beyond the scope of this article.

Routing in Directional Networks: Assuming the device and network parameters (n, r, α) are selected for a high likelihood of connectivity, routing protocols can be established for such directional networks. In contrast to traditional ad hoc routing protocols based on reverse path routing, the directional links necessitate that forward and reverse paths between network nodes often be distinct. A circuit-based paradigm, as illustrated in **Figure 8** (where the blue entity is the network sink), must be employed to facilitate bidirectional communications amongst network nodes that primarily transmit via unidirectional links. Here, circuits or loops are the fundamental entity for routing that guarantees one node can communicate to and from another node or network sink. All circuits including the network sink represent an uplink and downlink path from a MMSN node to the sink.

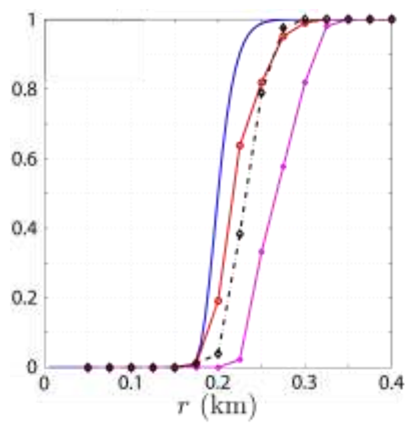


Figure 5. Probability of connectivity and no isolated node for beam width 40 degrees

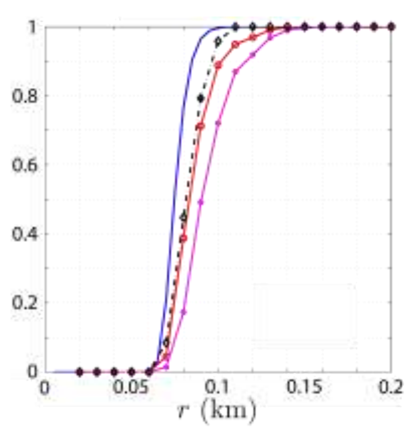


Figure 6. Probability of connectivity and no isolated node for beam width 270 degrees

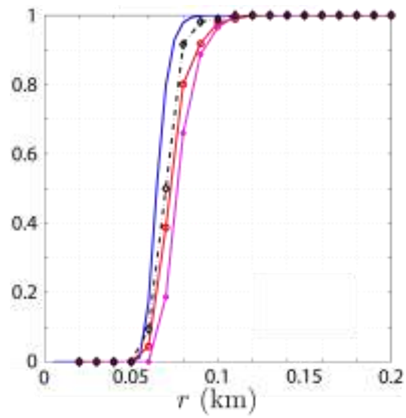


Figure 7. Probability of connectivity and no isolated node for beam width 360 degrees

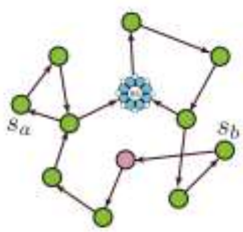


Figure 8. Circuit-based routing facilitates bidirecitonal links

For parameters (n, r, α) that guarantee a high likelihood of connectivity, it can be empirically and analytically shown that the number of hops in a circuit will usually not exceed 6 nodes making such a paradigm for directional link network routing potentially feasible. Topology discovery and route rediscovery mechanisms must account for the asymmetry in uplink and downlink routing, which naturally creates overhead. However, as we discuss next, this asymmetry can aid in network routing security.

Network Security in Directional Networks: Given the application space of MMSNs, security is of fundamental importance. Directional communications naturally lends itself to a more secure solution at the physical layer due to the more limited size of the communication footprint (see **Figure 1**), which makes interception of a communication beam more difficult. However, questions naturally arise as to whether there are any higher-level network security benefits of directional transmission paradigms.

The conventional threat model for ad hoc and sensor networks includes a high likelihood of insider attack. Thus, any network entity (excluding the sink) can potentially become corrupt. Given the high degree of coordination for such tasks as routing, even the corruption of a single node may have significant effects. A common strategy of legitimate network nodes is therefore to avoid collaboration with potentially corrupt nodes; thus, identification of such nodes is essential.

We assert in this paper that the asymmetry in communications warranted by directional link networks makes the network more secure. First, if traditional mechanisms to ensure successful data delivery are employed (e.g., via the use of ACK packets), a corrupt node in an uplink path would not be able to influence an ACK coming through a downlink path, thus alerting the network of a potential problem. Furthermore, for an attacker to hide such unwanted behavior, it would have to influence both the uplink and downlink paths thus raising the difficulty of the attack. For example, an attacker would have to corrupt two nodes in appropriate positions (depending on the topology) of the network.

Standard routing attacks geared for reverse path routing mechanisms no longer apply to a circuit-based approaches also providing inherent protection against naïve hackers. Future research efforts of the author and her group involve quantitative assessment of the trade-off between connectivity and security of directional link MMSNs.

Final Remarks: As MMSN systems emerge, we are at an exciting phase of development in which novel devices for sensing, communications and actuation must be employed. One class of such devices makes use of directional link communications to facilitate high-speed communications at lower power consumption. This article introduced some interesting aspects of directional link networking research and highlighted emerging challenges.

Useful Links:

1. Wireless Optical Sensor Networks: Connectivity, Routing and Security (Publications): http://www.ece.tamu.edu/~deepa/pub.html#wosn
2. Directional RF Sensor Networks: Connectivity and Security (Publications):

http://www.ece.tamu.edu/~deepa/pub.html#dirrf

Deepa Kundur received the B.A.Sc., M.A.Sc., and Ph.D. degrees all in Electrical and Computer Engineering in 1993, 1995, and 1999, respectively, from the University of Toronto, Canada. In January 2003, she joined the Department of Electrical Engineering at Texas A&M University, College Station, where she is a member of the Wireless Communications Laboratory and holds the position of Associate Professor. Before joining Texas A&M, she was an Assistant Professor at the Edward S. Rogers Sr. Department of Electrical and Computer Engineering at the University of Toronto where she was the Bell Canada Junior Chair-holder in Multimedia and an Associate Member of the Nortel Institute for Telecommunications.

Dr. Kundur's research interests include protection of scalar and broadband sensor networks, multimedia security, and computer forensics. She is an elected member of the IEEE Information Forensics and Security Technical Committee, vice-chair of the Security Interest Group of the IEEE Multimedia Communications Technical Committee and on the editorial boards of the IEEE Transactions on Multimedia and the EURASIP Journal on Information Security. More recently, she has been a guest editor for the 2007 EURASIP Journal on Advances in Signal Processing Special Issue on Visual Sensor Networks and the 2009 EURASIP Journal on Information Security Special Issue on Enhancing Privacy Protection in Multimedia Systems. She has been the recipient of the 2005 Tenneco Meritorious Teaching award, the 2006 Association of Former Students College Level Teaching award, and the 2007 Outstanding Professor Award in the ECE Department.

IG Corner

Home Networking, Shaping up the Pacemaker of the Digital Future

A brief introduction to the Home Network Interest Group (HNIG)

Madjid Merabti, Liverpool John Moores University, UK
Heather Yu, Huawei Technology, USA
M.Merabti@ljmu.ac.uk, heathery@ieee.org

The concept of networked home is not new. At first, home network is merely about connecting multiple computers and accessories such as printers within the home and to the Internet for shared computing resources and Internet access. Today, a typical home environment may contain a few networked devices such as personal computers (PCs), cellular phones, and web cameras or surveillance systems. Having the ability to access YouTube video and playback a video stored on your computer right on your TV screen is a dream just realized by several Consumer Electronic (CE) companies with their latest TV releases. And the future is even more exciting. It has long been envisioned that future home environment will have a vast array of Internet-connected devices that are far beyond the 3-screens (PC, TV, and cell phone). Networked appliances, sensors, as well as embedded systems will be connected through a self configurable, manageable, and secure home network and accessible through the Internet at anywhere and any time. While on one hand this gives users the freedom to access a rich variety of devices and services in the vicinity of their homes, on the other it increases the depth of knowledge required to understand what is happening in such a complex networking environment. Having the ability to check the inventory of your refrigerator while shopping in a grocery store, turn on the coffee maker when driving back home, watching your little girl's recital video with your family members and friends in different locations and simultaneously, and check out various functions of your house even from thousands of miles away requires new home networking infrastructure with intelligent service provision platform, interoperable network standards, comprehensive networking and knowledge processing technologies, and methods for appropriate information security and access control. This leads to the demand for novel technologies that offer the ability to connect, manage, control, and protect each and every home networking element, to represent and organize the diverse set of information effectively, and to process information intelligently for transparent and non-hassle accessibility.

Established in 2004 as one of the first interest groups within MMTC, HNIG was formed to provide a platform for its members, and the home networking and communications research, development, and standardization community at large, to interact and exchange technical ideas, to identify major R&D challenges, and to collaborate and investigate solutions in the development of consumer communications and home networking technologies. The founding of this IG has offered a fertile ground for MMTC members to join the effort in promoting the advancement of this technology area. For instance, one of the major conferences in the field, Consumer Communications and Networking Conference, is sponsored by the HNIG and many of our MMTC members are highly involved or leading the organization of the conference over the years. Another key activity supported by the HNIG is the Consumer Communications and Networking Series in the Communications Magazine where the HNIG Chair, Madjid Meratbi, is one of the co-editors together with Mario Kolberg, and MMTC past Chair Stanley Moyer. The large amount of submissions to this series evidences the research effort in the field. Besides these key activities, HNIG has also sponsored journal special issues and conference symposia, workshops, and special sessions.

The IG is open to everyone. We sincerely welcome new members who are interested in the field to join the IG and actively involved in the activities. Together we can help the advancement of the field and make a difference.

Madjid Merabti is Director of the School of Computing & Mathematical Sciences, Liverpool John Moores University, UK. A graduate of Lancaster University, he has over 20 years experience in conducting research and teaching in Distributed Multimedia Systems (Networks, Operating Systems, and Computer Security). Madjid leads the Distributed Multimedia Systems and Security Research Group. He is a member and chair of a number of conference TPCs and is chair of the PostGraduate Networking Symposium series (PGNet 2006) for UK PhD students. He was program chair for DRM '06, CCNC.

Heather Yu is a senior manager and head of the Multimedia Content Networking research team at Huawai Technologies (USA) She received her Ph.D. in Electrical Engineering from Princeton University. Currently she is serving as Associate Editor-in-Chief of the Journal of Peer-to-Peer Networking and Applications, Chair of the new ComSoc technical subcommittee on Human Centric Communications, a voting member of the GLOBECOM/ICC Technical Content Committee, and a member of ComSoc's Strategic Planning Committee. Her research interests include multimedia communications and multimedia content access and delivery. She has published 2 books, more than 60 technical papers and holds 23 U.S. patents.

MMTC COMMUNICATIONS & EVENTS

Call for Papers of Selected Journal Special Issues

Journal of Communications
Special Issue on Multimedia Computing and Communications

Guest Editors: Fan Zhai, Homer Chen, Thomas Stockhammer, Touradj Ebrahimi
Paper Submission deadline: **September 1**, 2009
Target Publishing Issue: 2nd Quarter, 2010
CfP Weblink: http://www.academypublisher.com/jcm/si/jcmsi_mcc.html

International Journal of Digital Multimedia Broadcasting
Special Issue on Video Analysis, Abstraction, and Retrieval: Techniques and Applications

Guest Editors: Jungong Han, Ling Shao, Peter H.N. de With, Ling Guan
Paper Submission deadline: **September 1**, 2009
Target Publishing Issue: March 1, 2010
Cfp Weblink: http://www.hindawi.com/journals/ijdmb/si/varta.html

Multimedia System Journal
Special Issue on Wireless Multimedia Transmission Technology and Application

Guest Editors: Gabriel-Miro Muntean, Pascal Frossard, Haohong Wang, Yan Zhang, Liang Zhou
Paper Submission deadline: **Jan. 15**, 2010
Target Publishing Issue: 4th Quarter, 2010
Cfp Weblink: http://www.ifi.uio.no/MMSJ/CFP_SI_Wireless_MMSJ.pdf

Call for Papers of Selected Conferences

ICC 2010

Website:	http://www.ieee-icc.org/2010/
Dates:	May 23-27, 2010
Location:	Cape Town, South Africa
Submission Due:	**Sept. 4, 2009**

S3DV 2009

Website:	http://crome.cs.ualberta.ca/S3DV2009/
Dates:	Sept. 27, 2009
Location:	Kyoto, Japan
Submission Due:	**July. 10, 2009**

iWAM 2009

Website:	http://crome.cs.ualberta.ca/iWAM/
Dates:	Oct. 24, 2009
Location:	Beijing, China
Submission Due:	**June 21, 2009**

Next Issue Partial Content Preview

Distinguished Position Paper Series:
A Framework on Multimodal Communications from Human Perspective
Zhengyou Zhang, Microsoft Research, USA

Trends in Mobile Graphics
Kari Pulli, Nokia Research, USA

3D Visual Quality
Wijnand IJsselsteijn, Eindhoven University of Technology, Sweden

3D Video Capturing and Analysis
Cha Zhang, Microsoft Research, USA

3D Visual Content Compression for Communications
Karsten Müller, Heinrich Hertz Institute (HHI), Germany

Interactive 3D Online Applications
Irene Cheng, University of Alberta, Canada

Focused Technology Advances Series: Robust Distributed Resource Allocation: A Marriage between Optimization Theory and Game Theory
Jiangwei Huang, Chinese University of Hong Kong, China

E-Letter Editorial Board

EDITOR-IN-CHIEF

Haohong Wang
TCL-Thomson Electronics
USA

EDITOR

Philippe Roose
IUT of Bayonne
France

Guan-Ming Su
Marvell Semiconductors
USA

Antonios Argyriou
Phillips Research
Netherlands

Chonggang Wang
NEC Laboratories America
USA

Shiguo Lian
France Telecom R&D Beijing
China

MMTC Officers

CHAIR

Qian Zhang
Hong Kong University of Science and Technology
China

VICE CHAIRS

Wenjun Zeng
University of Missouri, Columbia
USA

Zhu Li
Hong Kong Polytechnic University
China

Madjid Merabti
Liverpool John Moores University
UK

Nelson Fonseca
Universidade Estadual de Campinas
Brazil

SECRETARY
Bin Wei
AT&T Labs Research
USA