Special Issue on Video Analysis on Resource-Limited Systems

Camera sensors, embedded in surveillance cameras, webcams, cameras in mobile devices, wearable cameras, and so on, are becoming pervasive in our society. Moreover, it is believed that even more cameras will be embedded everywhere as smart sensors in the future, driven by the decreasing cost and important applications in safety and security, smart environments, gaming and entertainment, and visual communication. To automatically analyze the massively increasing visual data, computer vision and video analysis techniques have received much attention in the recent years.

In many real-world video analysis applications, the available resources are limited. This could mean low-quality data (e.g., limited imaging resolution/sensor size/frame rate), such as video footage from surveillance cameras and videos captured by consumers via mobile or wearable cameras. Another dimension comes from limited amount of processing power, for example, on mobile phones. The low-quality data could be caused by the poor sensor performance, simple optics, motion blur, or environmental factors such as illumination. Due to the limitations on video storage and transmission, the captured videos are often compressed, which may also result in low-quality video data. The sensors in the nonvisible spectrum (near infrared, far infrared, and so on) generate low-resolution videos with much noise, too. On the other hand, the processing power on the mobile or wearable devices is still limited for traditional video analysis tasks. There is a huge demand for video analysis and computer vision techniques on resource-limited systems, which could enable many practical applications, for example, face identification for law enforcement, abnormal behavior detection for security, place/sign recognition for mapping and localization, augmented reality on mobile phones, to name a few. However, video analysis on resource-limited systems is still an under-explored field. The existing video analysis research mainly focuses on high-performance systems, where high-quality video data or powerful computing platforms are considered.

There are many challenges when addressing video analysis on resource-limited systems. For example, how to effectively extract representative visual features from low-quality data? How to fuse multiple low-resolution frames for reliable recognition? How to accelerate vision algorithms for use on mobile platforms? How to mitigate degrading factors caused by the low-quality data? We have to adapt the existing techniques developed for high-performance systems or find new approaches suitable for resource-limited systems. This special issue seeks to present and highlight the recent developments in the area of video analysis on resource-limited systems.

Date of current version October 5, 2011.
Digital Object Identifier 10.1109/TCSVT.2011.2165795

Our call for papers was enthusiastically greeted by the research community, and we received over 60 submissions from both academia and industry. These submissions cover a wide range of topics related to video analysis on resource-limited systems. The corresponding authors represent broad diversity in geographic location, including North America, Europe, and Asia. The quite large number of submissions posed a challenge for reviewing. We sought assistance for reviewing these papers from more than 200 reviewers. Each submission was reviewed by at least three experts, to ensure high reviewing standards. Based on the reviewers' recommendations and after discussion among the Guest Editors and Editor-in-Chief, 17 papers were eventually accepted for the special issue. The papers included can be categorized as follows.

I. Low-Quality Data Enhancement

The video data acquired by the widely used inexpensive surveillance cameras is of low quality, i.e., with limited resolution and frame rate. This makes it difficult to perform high-level vision tasks, e.g., face image analysis. Super-resolution (SR) is one of mechanisms that enhance the imaging resolution, which obtains a high-resolution (HR) image from one or more low-resolution (LR) images. In the paper entitled "Extracting a Good Quality Frontal Face Image from a Low-Resolution Video Sequence," Nasrollahi and Moeslund investigate face super-resolution for face recognition in surveillance videos. They use a three-step approach to produce a face-log containing similar frontal face images. By combining reconstruction-based and learning-based SR algorithms, they obtain the resolution improvement with a factor of four. In another paper entitled "Fast Facial Image Super-Resolution via Local Linear Transformations for Resource-Limited Applications" by Huang and Wu, the problem of reconstructing a HR face image from a single LR face image is addressed. They utilize multiple local linear transformations to approximate the nonlinear mapping between LR and HR face images, represented with orthogonal matrices derived by Procrustes analysis.

In practice the surveillance videos are compressed before transmission. Standard compression on surveillance videos can reduce the accuracy of followup video analysis (e.g., object tracking). In their paper "Low-Complexity Tracking-Aware H.264 Video Compression for Transportation Surveillance," Soyak *et al.* propose a tracking-aware video compression algorithm by utilizing tracking accuracy as compression criterion in lieu of mean squared error metrics. Their application-specific compression can significantly reduce bitrate while maintaining comparable tracking accuracy.

In many cases (e.g., hand-held cameras), video quality is degraded due to unwanted camera movements, which blur and introduce disturbing jerkiness in the recorded videos.

The paper by Puglisi and Battiato, entitled "A Robust Image Alignment Algorithm for Video Stabilization Purposes," presents a fast image alignment algorithm for video stabilization, where a block based local motion estimator is combined with voting based alignment. In another paper entitled "System-on-Chip Solution of Video Stabilization for CMOS Image Sensors in Hand-Held Devices," Kim *et al.* address video stabilization at the sensor level for handheld devices. Considering most of CMOS sensors in handheld devices utilize a rolling shutter to increase sensitivity, which causes severe distortions, they model the CMOS sensor with a section-wise CCD sensor array, and propose a CMOS seven parameter model. To accelerate video stabilization, they present cache based optimization techniques and the incremental form of seven-parameter model to reduce resource consumption.

II. Camera Sensing and Processing

Video capturing and processing is usually based on a sequence of images. In the paper entitled "Advantages of Selective Change Driven Vision for Resource-Limited Systems," Pardo *et al.* propose the selective change driven (SCD) vision strategy, which delivers only the pixels that have undergone the greatest change in illumination since the last time they were readout. SCD vision processes a flow of pixels, instead of the image flow, leading to far lower bandwidth and resource requirements. They present a CMOS sensor using this strategy, along with a resource-limited system implementing object tracking.

A visual sensor network (VSN) consists of a large number of camera nodes (integrating image sensor, embedded processor, and wireless transceiver) with tight resource limitations. In the paper entitled "Resource-Aware Coverage and Task Assignment in Visual Sensor Networks," Dieber *et al.* focus on finding optimal VSN configurations, which can select cameras for sufficient monitoring, set their frame rate and resolution to fulfill the quality of service requirements, and assign processing tasks to the cameras. They formulate the configuration problem and describe an efficient approximation based on an evolutionary algorithm. To increase energy efficiency and prolong battery-life of battery-powered camera node, in their paper "Adaptive Methodologies for Energy-Efficient Object Detection and Tracking with Battery-Powered Embedded Smart Cameras," Casares and Velipasalar propose self-adapting methodologies, which set the camera microprocessor to the idle state even when there are tracked objects in the scene. They also present a feedback method for efficient object detection and tracking.

High performance and low-power very large scale integrations are required to implement complex media processing applications on mobile devices, where heterogeneous multicore processors are a promising solution. To remove data transfer bottlenecks in such platforms, in the paper entitled "Memory Allocation Exploiting Temporal Locality for Reducing Data-Transfer Bottlenecks in Heterogeneous Multicore Processors," Waidyasooriya *et al.* propose a memory allocation method that exploits the temporal and spatial locality of memory access in media processing applications. Evaluation results show their method reduces the total processing time significantly compared to the previous works.

III. Video Analysis on Mobile Devices

With mobile devices (equipped with a camera) become more prevalent, video analysis on mobile devices has become a hot topic, with many interesting applications. Mobile platforms normally have limited computational power and memory, and the applications require fast response. Therefore, many efforts have been devoted to accelerating existing video analysis algorithms for mobile platforms. In the paper entitled "Adaptive Sampling for Feature Detection, Tracking and Recognition on Mobile Platforms" by Ebrahimi and Mayol-Cuevas, adaptive sampling is used to speed up local feature detection and recognition for object recognition on mobile platforms. Their method is implemented on a chipset commonly found on smartphones, and is shown to improve both the processing time and memory footprint. The paper entitled "Integrated Content and Context Analysis for Mobile Landmark Recognition" by Chen *et al.* introduce an effective method that integrates content and context analysis for landmark recognition on mobile devices. Based on spatial pyramid decomposition and cell saliency estimation, they develop a new bags-of-words for efficient content analysis.

Augmented reality (AR) on smartphones is an interesting application. In their paper, "Video-Based In Situ Tagging on Mobile Phones," Lee *et al.* consider a mobile AR tagging application, where a real-world scene is augmented by virtual contents through a mobile phone camera. They adapt the perspective patch recognition technique for mobile phones, and exploit the phone accelerometers to reduce user intervention. The paper by Roters *et al.*, entitled "Recognition of Traffic Lights in Live Video Streams on Mobile Devices," presents a mobile computer vision system that helps visually impaired pedestrians cross roads by detecting and recognizing traffic lights. Taking into account the limited resource on mobile devices, their system consists of four steps: localization, classification, video analysis, and time-based verification. The temporal analysis alleviates the inherent problems such as occlusions and falsified colors. A prototype system implemented on Nokia N95 is tested in real environment for detecting German traffic lights.

Video telephony based on 3G-324M is becoming popular due to prevalence of smartphones. Media synchronization is an issue to be solved in order to achieve satisfactory video communication quality. The paper entitled "Synchronization Quality Enhancement in 3G-324M Video Telephony" by Jang and Na analyzes existing solutions and proposes an enhanced mobile media synchronization scheme. Their scheme consists of dynamic multiplex code selection and a frame-to-frame adaptive playout strategy. Implemented and evaluated in a real 3G mobile phone, the proposed scheme is shown to decrease the audio-visual synchronization inconsistency by more than 10% compared to a conventional scheme.

IV. Other Resource-Limited Platforms

Visual surveillance from low-altitude airborne platforms plays a role in urban planning and traffic monitoring. In the paper entitled "Vehicle Detection and Motion Analysis in Low-Altitude Airborne Video Under Urban Environment," Cao *et al.* present a computationally efficient framework for moving vehicle detection and motion analysis in low-altitude airborne videos. They introduce a boosting light and pyramid sampling histogram of oriented gradients feature extraction step, followed by a spatio-temporal appearance-related similarity step that correlates detections across frames to estimate the trajectories of the vehicles.

Wireless video capsule endoscopy is a state-of-the-art technology to acquire video of human intestine for medical diagnostics. The main challenge in video capsule endoscopic systems is to reduce the area and power consumption while maintaining acceptable video reconstruction. The paper by Khan and Wahid, entitled "Low Power and Low Complexity Compressor for Video Capsule Endoscopy," presents a low-complexity image compression algorithm, by exploiting the properties of endoscopic images. The proposed compressor eliminates the need of memory buffer and temporary storage.

Existing stereo vision algorithms have focused on the accuracy, to minimize the error between the generated disparity map and the ground truth. The practical issues including computational complexity and resource consumption are neglected in the literature. The paper entitled "Dense Disparity Real-Time Stereo Vision Algorithm for Resource Limited Systems" by Tippetts *et al.* investigates stereo vision algorithms that trade off the accuracy for resource reduction. They introduce an intensity profile shape matching algorithm capable of producing dense disparity maps in real time on resource-limited systems, which provides enough accuracy for some practical applications.

We believe the papers selected for publication in this special issue capture the state-of-the-art research in the area of video analysis on resource-limited systems. We hope you will enjoy this collection of papers and this special issue can spark and stimulate further research and development in this area.

ACKNOWLEDGMENT

We have many people to thank for their contributions to this special issue. We would like to thank all of the authors for submitting their papers to this special issue. We are also very grateful to all the reviewers who provided the authors with timely, critical, yet constructive feedback; their evaluations helped shape this special issue. Finally, our special thanks go to the Editor-in-Chief, Dr. H. Gharavi, for his constant support and guidance throughout the entire process that has led to the publication of this special issue.

RAMA CHELLAPPA, *Fellow, IEEE*
University of Maryland
College Park, MD 20742 USA

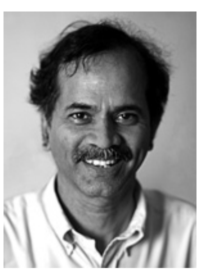
ANDREA CAVALLARO
Queen Mary University of London
London E1 4NS, U.K.

YING WU, *Senior Member, IEEE*
Department of Electrical and Computer Engineering
Northwestern University
Evanston, IL 60208 USA

CAIFENG SHAN, *Member, IEEE*
Philips Research
Eindhoven 5656, The Netherlands

YUN FU, *Senior Member, IEEE*
Department of Computer Science and Engineering
State University of New York at Buffalo
Buffalo, NY 10018 USA

KARI PULLI
NVIDIA Research
Santa Clara, CA 95050 USA



Rama Chellappa (F'92) received the B.E. (hons.) degree from the University of Madras, Chennai, India, in 1975, the M.E. (distinction) degree from the Indian Institute of Science, Bangalore, India, in 1977, and the M.S.E.E. and Ph.D. degrees in electrical engineering from Purdue University, West Lafayette, IN, in 1978 and 1981, respectively.

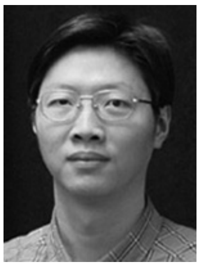
Since 1991, he has been a Professor of electrical engineering and an Affiliate Professor of computer science with the University of Maryland, College Park. He is also affiliated as the Director with the Center for Automation Research and is a Permanent Member with the Institute for Advanced Computer Studies. Currently, he holds a Minta Martin Professorship with the College of Engineering. Prior to joining the University of Maryland, he was an Assistant Professor from 1981 to 1986, an Associate Professor from 1986 to 1991, and the Director of the Signal and Image Processing Institute, University of Southern California, Los Angeles, from 1988 to 1990. Over the last 30 years, he has published numerous book chapters, peer-reviewed journals, and conference papers in image and video processing, analysis, and recognition. He has also co-edited/co-authored six books on neural networks, Markov random fields, face/gait-based human identification, and activity modeling. His current research interests include face and gait analysis, 3-D modeling from video, automatic target recognition from stationary and moving platforms, surveillance and monitoring, hyper spectral processing, image understanding, and commercial applications of image processing and understanding.

Dr. Chellappa has served as an Associate Editor of four IEEE transactions. He was a Co-Editor-in-Chief of *Graphical Models and Image Processing* and also served as the Editor-in-Chief of the IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE from 2001 to 2004. He served as a member of the IEEE Signal Processing Society Board of Governors from 1996 to 1999 and as its Vice President of Awards and Membership from 2002 to 2004. He completed a two-year term as the President of the IEEE Biometrics Council. He has received several awards, including the NSF Presidential Young Investigator Award, four IBM Faculty Development Awards, an Excellence in Teaching Award from the School of Engineering at USC, and two Paper Awards from the International Association of Pattern Recognition. He received the Society, Technical Achievement and Meritorious Service Awards from the IEEE Signal Processing Society and served as its Distinguished Lecturer. He also received the Technical Achievement and Meritorious Service Awards from the IEEE Computer Society. At the University of Maryland, he was elected a Distinguished Faculty Research Fellow, as a Distinguished Scholar-Teacher (both university wide awards), received the Outstanding Faculty Research Award and the Poole and Kent Teaching Award for the Senior Faculty from the College of Engineering, an Outstanding Innovator Award from the Office of Technology Commercialization, and an Outstanding GEMSTONE Mentor Award. In 2010, he was recognized as an Outstanding ECE by Purdue University, West Lafayette, IN. He is a fellow of the International Association for Pattern Recognition and the Optical Society of America.

Andrea Cavallaro received the Ph.D. degree in electrical engineering from the Swiss Federal Institute of Technology, Lausanne, Switzerland, in 2002, and the Laurea (summa cum laude) degree in electrical engineering from the University of Trieste, Trieste, Italy, in 1996.

He was a Research Fellow with British Telecommunications, Adastral Park, Suffolk, U.K., from 2004 to 2005. He has published more than 100 journal and conference papers, and two books *Multi-Camera Networks: Principles and Applications* (Elsevier, 2009) and *Video Tracking: Theory and Practice* (Wiley, 2011). He is currently a Professor of multimedia signal processing with Queen Mary University of London, London.

Dr. Cavallaro was awarded the Royal Academy of Engineering Teaching Prize in 2007, three student paper awards on target tracking and perceptually sensitive coding at IEEE ICASSP in 2005, 2007, and 2009, and the Best Paper Award at IEEE AVSS 2009. He is an Associate Editor for the IEEE Signal Processing Magazine, the IEEE Transactions on Signal Processing, and the *Journal on Information Security*. He is an Elected Member of the IEEE Signal Processing Society, Image, Video, and the Multidimensional Signal Processing Technical Committee. He has also served as an Elected Member of the IEEE Signal Processing Society, the Multimedia Signal Processing Technical Committee, as an Associate Editor for the IEEE Transactions on Multimedia, and as a Guest Editor for several journals including *Computer Vision and Image Understanding* (Elsevier), the *International Journal of Computer Vision*, the IEEE Signal Processing Magazine, and the IEEE Transactions on Circuits and Systems for Video Technology. He was the General Chair for IEEE/ACM ICDSC 2009, BMVC 2009, M2SFA2 2008, SSPE 2007, and IEEE AVSS 2007. He is the Technical Program Chair of IEEE AVSS 2011 and was the Technical Chair of the European Signal Processing Conference (EUSIPCO 2008) and of WIAMIS in 2010.



Ying Wu (SM'06) received the B.S. degree from the Huazhong University of Science and Technology, Wuhan, China, in 1994, the M.S. degree from Tsinghua University, Beijing, China, in 1997, and the Ph.D. degree in electrical and computer engineering from the University of Illinois at Urbana-Champaign (UIUC), Urbana, in 2001.

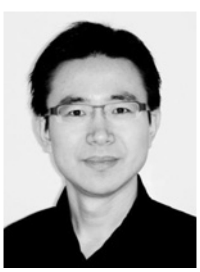
From 1997 to 2001, he was a Research Assistant with the Beckman Institute for Advanced Science and Technology, UIUC. During 1999 and 2000, he was a Research Intern with Microsoft Research, Redmond, WA. In 2001, he joined the Department of Electrical and Computer Engineering, Northwestern University, Evanston, IL, as an Assistant Professor. He is currently an Associate Professor of electrical engineering and computer science with Northwestern University. His current research interests include computer vision, image and video analysis, pattern recognition, machine learning, multimedia data mining, and human–computer interaction.

Dr. Wu serves as an Associate Editor for the IEEE Transactions on Image Processing, the *SPIE Journal of Electronic Imaging*, and the *IAPR Journal of Machine Vision and Applications*. He received the Robert T. Chien Award at UIUC in 2001, and the NSF CAREER Award in 2003.

Caifeng Shan (S'05–M'08) received the B.Eng. degree in computer science from the University of Science and Technology of China, Hefei, China, in 2001, the M.Eng. degree in pattern recognition and intelligent system from the Institute of Automation, Chinese Academy of Sciences, Beijing, China, in 2004, and the Ph.D. degree in computer vision from the Queen Mary University of London, London, U.K., in 2007.

He is currently a Senior Scientist with Philips Research, Eindhoven, The Netherlands. He has authored about 50 technical papers and six patent applications. He edited two books *Video Search and Mining* (Berlin, Germany: Springer, 2010) and *Multimedia Interaction and Intelligent User Interfaces: Principles, Methods and Applications* (Springer, 2010). His current research interests include computer vision, pattern recognition, image/video processing and analysis, machine learning, multimedia, and related applications.

Dr. Shan has been the Guest Editor of the IEEE Transactions on Multimedia and *Signal Processing (Elsevier)*. He organized several international workshops at flagship conferences such as IEEE ICCV and ACM Multimedia. He has served as a Program Committee Member and a Reviewer for numerous international conferences and journals. He was the recipient of the 2007 Chinese Government Award for Outstanding Self-Financed Students Abroad.



Yun Fu (S'07–M'08–SM'11) received the B.Eng. degree in information engineering in 2001 and the M.Eng. degree in pattern recognition and intelligence systems in 2004, both from Xi'an Jiaotong University, Xi'an, China, and the M.S. degree in statistics in 2007 and the Ph.D. degree in electrical and computer engineering in 2008, both from the University of Illinois at Urbana-Champaign, Urbana.

He was a Research Intern with Mitsubishi Electric Research Laboratories, Cambridge, MA, in 2005, and with the Multimedia Research Lab of Motorola Labs, Schaumburg, IL, in 2006. He joined BBN Technologies, Cambridge, as a Scientist in 2008. He was a part-time Lecturer with the Department of Computer Science, Tufts University, Medford, MA, in 2009. He joined the Department of Computer Science and Engineering, State University of New York, Buffalo, as an Assistant Professor in 2010. His current research interests include applied machine learning, social media analytics, human-centered computing, pattern recognition, and intelligent vision systems.

Dr. Fu was the recipient of the 2002 Rockwell Automation Master of Science Award, the Edison Cups of the 2002 GE Fund Edison Cup Technology Innovation Competition, the 2003 Hewlett-Packard Silver Medal and Science Scholarship, the 2007 Chinese Government Award for Outstanding Self-Financed Students Abroad, the IEEE ICIP 2007 Best Paper Award, the 2007–2008 Beckman Graduate Fellowship, the 2008 M. E. Van Valkenburg Graduate Research Award, the ITESOFT Best Paper Award of the 2010 IAPR International Conference on the Frontiers of Handwriting Recognition, the 2010 Google Faculty Research Award, the IEEE ICME 2011 Quality Reviewer, and the 2011 IC Post-Doctoral Research Fellowship Award. He is the Associate Editor for the IEEE Transactions on Circuits and Systems for Video Technology. He is a Life Member of ACM, SPIE, the Institute of Mathematical Statistics, and a 2007–2008 Beckman Graduate Fellow.

Kari Pulli received the B.Comp.Sci. degree from the University of Minnesota, Minneapolis, in 1989, the M.Sc. and Lic.Tech. degrees in computer engineering from the University of Oulu, Oulu, Finland, in 1991 and 1993, respectively, the Ph.D. degree in computer science from the University of Washington, Seattle, in 1997, and the MBA degree from the University of Oulu in 2001.

He was a Research Associate with Stanford University, Stanford, CA, where he worked on the Digital Michelangelo Project from 1998 to 1999. In 1999, he joined Nokia, Oulu, leading much of its graphics research, technology, and standardization (e.g., OpenGL ES, M3G, OpenVG). He taught graphics as a Docent (Adjunct Faculty) in Oulu until 2004. He was a Visiting Scientist with the Massachusetts Institute of Technology, Cambridge, from 2004 to 2006. He headed a research team with the Nokia Research Center, Palo Alto, CA, from 2006 to 2011, working on mobile imaging, especially on computational photography and mobile augmented reality. He was nominated a Nokia Fellow and was a member of the CEO's Technology Council. In 2011, he joined NVIDIA Research, Santa Clara, CA, as a Senior Director heading the Mobile Visual Computing Team. He has written a book on mobile 3-D graphics (San Francisco, CA: Morgan Kaufmann, 2007), has given numerous conference keynotes, and is a member of editorial boards at *Computers and Graphics* and the IEEE Computer Graphics and Applications.