Fatih Porikli, François Brémond,
Shiloh L. Dockstader, James Ferryman,
Anthony Hoogs, Brian C. Lovell,
Sharath Pankanti, Bernhard Rinner,
Peter Tu, and Péter L. Venetianer

Video Surveillance: Past, Present, and Now the Future

Video surveillance is a part of our daily life, even though we may not necessarily realize it. We might be monitored on the street, on highways, at ATMs, in public transportation vehicles, inside private and public buildings, in the elevators, in front of our television screens, next to our baby's cribs, and any spot one can set a camera.

It was around for almost a century. As this 1939 *Popular Science* news article cleverly mentions, there is no arguing with the testimony of the movie camera: "Mounted on the dashboard of his patrol car, with its lens pointing forward through the windshield, a motion-picture camera belonging to Officer R. Galbraith of the California Highway Patrol takes photographs of the automobiles he trails along the highways, making a permanent film record of any traffic violations for possible later use in court" [1].

Surely, automated video surveillance would follow soon. A 1963 U.S. patent described a television system for detection of differences by "determining changes of interest in the scene and ignoring other changes by comparing digitized image point averages over prior scans of the same sample points" [2]. Another U.S. patent granted in 1966 explains a television surveillance system that outputs an alarm by using the differences between the sample data and data average [3].

Video surveillance offered an enticing hope as encapsulated in this 1965 *Scientific Mechanics* news article: "Here is a modern solution to the problem that is plaguing every large city in America today! The police do their best, but they cannot be everywhere at once in a verdant park. They need help in the form of a surveillance system of some kind. And it isn't necessary to dream up a Buck Rogers-type 'seeing eye.' The system already exists; it is called closed-circuit television, or CCTV" [4].

Like these exciting news articles, several studies in the past forecasted that automated video surveillance was poised for an explosive growth thanks to dropping camera prices, increasing network connectivity, and the massive deployments taking place in North America, China, and Europe. Despite recent economic woes, this migration was expected to accelerate by adoption in a variety of vertical markets such as retail, banking, transportation, and education where it is not always about security.

So, what did happen? Where are we now?

This *IEEE Signal Processing Magazine* forum discusses the latest advances, challenges, and future of video surveillance. The invited forum members, who bring their expert insights, are François Brémond (INRIA Sophia Antipolis), Shiloh L. Dockstader (ITT Exelis), Anthony Hoogs (Kitware), James Ferryman (University of Reading), Brian C. Lovell (The University of Queensland), Sharath Pankanti (IBM T.J. Watson Research Center), Bernhard Rinner (Alpen-Adria-Universität Klagenfurt), Peter Tu (General Electric Global Research), and Péter L. Venetianer (ObjectVideo). The moderator of this forum is Fatih Porikli (MERL).

Our readers may agree or disagree with the ideas discussed next. In either case, we invite you to share your comments with us by e-mailing fatihporikli@ieee.org or spm.columns.forums@gmail.com.

Moderator: Would you please describe the promise of video surveillance? What were the premises to keep its promise?

Péter L. Venetianer: The key high-level promise of automated video surveillance is to have computers replace human eyes and perform the basic tasks of detecting events of interest better, cheaper, and/or more reliably.

Peter Tu: With the multitude of cameras that have been installed throughout the world, a significant number of these video streams would be channeled into robust video analytics devices capable of producing reliable and valuable metadata. The availability of such data would then result in kind of situational awareness that would endow various stakeholders with the ability to make truly informed decisions.

James Ferryman: The promise of digital video surveillance has been for fully (or quasi-) automated monitoring of indoor and outdoor scenes, in particular for alerting operators within a control room environment to events or activities of interest. Where large geographic areas are monitored by digital networks of tens, hundreds and occasionally thousands of closed-circuit television (CCTV) cameras, it is recognized that operators tasked to monitor such cameras suffer from information overload and short attention span [5]. The main premise for some degree of automated CCTV analysis is that it is inconceivable that all CCTV can be effectively monitored manually.

Bernhard Rinner: Video surveillance has a rather long history—at least in terms of computing. It started more than 50 years ago, when the first CCTV systems had been deployed to monitor specific sites from remote. During the 1980s, video surveillance began to spread, specifically targeting crime

Digital Object Identifier 10.1109/MSP.2013.2241312
Date of publication: 5 April 2013

1053-5888/13/$31.00©2013IEEE

prevention in public areas. In the 1990s, digitalization and the introduction of video analytics transferred CCTV from sole video transmission (and archiving) to distributed camera systems capable of performing low-level analysis in real time [6]. Now, video surveillance has become ubiquitous with a huge number of globally installed cameras and applications going well beyond safety including transportation, entertainment, and assisted living.

The promise and expectations of video surveillance have naturally changed over this long period. The main objective in the early days was simply to extend the visual sensing capabilities of the observer to the site of interest, while recent surveillance systems are expected to perform complex analysis tasks with the goal to understand what is going on in the monitored area.

Brian C. Lovell: There is no question that video surveillance has been an immensely successful technology. This is corroborated by the increasing numbers of cameras being installed worldwide and the sheer pervasiveness of the technology. Much of the focus and public discussion on video surveillance centers on the use of recorded surveillance footage to solve high-profile crimes. Whenever this happens, there are renewed calls for major investments in public surveillance networks. However, it is worth noting that the majority of CCTV cameras are usually owned privately and the primary reason for installation may have little to do with crime prevention.

Ideally, we would like to search videos with computers to detect events of interest, but with rare notable exceptions, the promise of video content analysis has not been realized. Instead, we are faced with the problem of human searching of the ever-increasing volumes of stored video. One solution being explored is crowd sourcing for content analysis.

Recently, we had the tragic murder of a woman who was abducted from a city street only meters from her home [7]. The CCTV camera that captured an image of her alleged attacker was a private CCTV camera recently installed in a bridal shop. Within days, the suspect was apprehended after the footage was placed on social media in a groundbreaking collaboration between the police and the public.

In the United Kingdom, social media is already being used to harness private CCTV to great effect. A U.K. police sponsored organization, "FaceWatch" [8] provides a service for private security footage to be uploaded onto a common online portal, so the public and businesses can help police solve the crime. For example, graffiti can be fairly easily recognized through comparing works and "tagging" of offenders and linking these together using geographic location. Many low-level crimes can thus be resolved in an extremely cost effective manner.

A common theme here is the linking of public and private surveillance networks to solve a great range of crimes more effectively. Indeed, In Hong Kong, for example, there is virtually no public CCTV surveillance network and yet Hong Kong is also one of the most surveilled societies, possibly even more so than the United Kingdom, where it is claimed that there is one CCTV camera for every 32 people [9]. Private CCTV footage is routinely used by the Hong Kong police to solve major crimes. An effective crime surveillance network is already present in most countries through private security cameras in bank autotellers, elevators, convenience and franchise stores, service stations, hotels, clubs, nursing and aged-care homes, and private residences, particularly large apartment blocks.

James Ferryman: I agree, applications for video surveillance vary widely. One of the main promises of video surveillance has been in the monitoring of public spaces such as schools, hospitals and sports grounds for personal safety, for example, volume crime prevention and detection. Other examples include the efficient management of transport networks, protection of (critical) infrastructure, border control, retail analytics, health care, and the reporting of sports statistics (e.g., ball possession in soccer).

Indeed, from a citizen's standpoint, one vision of surveillance could be stated as follows: In the context of public transport, the daily life of a citizen involves passing between different public environments. Such people spaces may be populated with multiple vision-based services offering a range of surveillance services that actively or passively support the activities of citizens within each environment, and promote the collective well being of these spaces.

For operators of critical infrastructure, video surveillance has been sold as a solution to robustly determine incursions and criminal intent. For police and law enforcement, video analytics has been sold as the answer to interrogating (e.g., searching for all instances of person X across Y cameras over time T) large volumes of video data. For the public at large, video surveillance has been sold through movies, such as *Minority Report*, and television drama series, such as the U.K. drama series *Spooks*, as a technology that is capable of respectively facilitating a natural and fluid pointing-gesture-controlled interface and the ability for intelligence operatives to recognize in real time, through automatic face recognition, individuals in a crowded street at a distance.

Sharath Pankanti: The fundamental assumption underlying typical video surveillance is that the visual information is the (only) predominant component for assessing situational awareness. That is, banks of displays conveying visual information about (remote) location is a practical surrogate for an alert living/sensing human monitoring the physical location. A typical municipal/retail command and control center will have camera monitors covering an entire wall and a bevy of humans monitoring all the incoming video feeds for suspicious activities. The remarkable growth of camera videos acquisition has lead to a situation wherein we have a shortage of personnel to monitor all of the data that is being generated. Such a scenario is typical in a video surveillance situation, where a massive number of cameras are being deployed to monitor large geographical areas, such as cities. Such human monitoring not only suffers from loss of attentiveness, since one cannot simultaneously focus on all the activities in all the cameras at once, but also from human fatigue and boredom while looking at these camera feeds for extended periods of time.

The premise underlying automated video surveillance is that we can relieve the perceptual overload on the operators monitoring the banks of displays by letting them adjudicate the events detected by video analytic systems in real time. In addition to real-time alerting capabilities, video analytic systems have also been used to enable the users to search for events of interest after the fact. The automated video analytic systems appear to be a reasonable practical approach to surveillance. Many commercial systems for intelligent urban surveillance exist in the market including the systems of companies. These systems have accomplished practically a lot in last two decades, especially, when we consider the capabilities of manual video surveillance systems or the cost of the "cop on the beat" approach.

For example, well-designed surveillance systems can handle hundreds to a few thousands of cameras covering a large metro city. There can be tens of thousands of events (vehicle traffic only) per day per camera covering a busy urban street. Therefore, the system can be handling on the order of trillion events a month! Modern surveillance systems can support indexing of such large amount of data with clever partitioning and federation strategies and can let the operators search and effectively navigate through these data. Such systems allow the user to automatically search for objects and can respond to requests such as "Show me all the people who entered this facility yesterday" or "Show me all the red cars that crossed this avenue last Sunday from 5 a.m. to 9 a.m." [10]. Many surveillance solutions can search the vehicles on license plate recognition or vehicle classification with appropriate high-resolution, high-speed camera coverage backed up by appropriate compute/network infrastructure.

Shiloh L. Dockstader: Initially, the promise of video surveillance was equivalent to offering users the ability to seamlessly and remotely perform effective intelligence, surveillance, and reconnaissance (ISR) operations. To some extent, this has been a well-achieved goal. Video cameras and surveillance infrastructures abound for countless applications ranging from those in defense to commercial and homeland security to home monitoring [11]. Furthermore, numerous commercial-off-the-shelf software systems, exploitation packages, and video camera options are available to support both ground and airborne video surveillance systems. From an interoperability perspective, a range of standards (e.g., MPEG-2, H.264, MJ2K, etc.) are available to guide the creation of standards-compliant video files and compression profiles. The Department of Defense/Intelligence Community Motion Imagery Standards Board focuses exclusively on furthering the state of the art in standards, interoperability, testing, and evaluation for video surveillance applications. One might argue that the promise of video surveillance was achieved decades ago….

Anthony Hoogs: Given the previous responses, I'll take a higher-level view of the question. The promise of automated video surveillance, or video analytics, is to make everyone safer, healthier, wealthier, and even happier. Why else would we allow the inherent, ever-increasing privacy intrusion so blatantly required for video surveillance? Why else would we sacrifice personal freedoms on such a grand scale? While few people realize the extent to which our privacy is already compromised, many appreciate the well-publicized successes of video surveillance, automated or otherwise, as Brian pointed out. At this point, most modern societies have decided that video surveillance is worth the cost. Improved safety is an obvious benefit, but improved health results from health-care monitoring applications, and in-home alerting for aging populations is emerging. Video surveillance in retail settings saves money by pinpointing perpetrators, and (probably) discouraging crime. These combined benefits make us happier as a society.

With increasing automation, all of these applications become more effective and more become possible. In the long run, the overhyped, unmet expectations of the video analytics industry, discussed at length in the next section, will not matter. When video analytics products are truly useful in their target domains, they will find a market regardless of whether experts predicted them to be viable years before they actually were.

Moderator: What are the currently confronted challenges then? Did video surveillance solutions meet the expectations?

Shiloh L. Dockstader: To a greater extent, however, video surveillance has implicitly promised much more than simply providing underlying hardware and software infrastructure and components. The initial promise of providing a remote, video-based ISR collection capability has actually been so successful that it has driven demand for video surveillance systems to insatiable levels. This has resulted in a derivative, much broader, and more comprehensive field of video surveillance and associated analytics, replete with new capabilities but also countless new expectations, issues, and problems. As such, there exists significant variation in the way end users characterize the terms "seamlessly" and "effective" as it pertains to ISR and video surveillance systems. From the perspectives of the broader community, video surveillance and video analytics are often seen as synonymous, where the promise (and potential) of video surveillance is now overwhelmingly unfulfilled.

Today's ubiquitous video collection devices and associated surveillance systems have created a big data problem in which far more data is collected than ever viewed or analyzed. The advanced analytics and processing and exploitation algorithms needed to distill this data into useful intelligence components still lacks in maturity. There are seemingly an infinite number of automated target detection, feature extraction, and tracking algorithms for video surveillance applications, yet not one that offers flawless performance. Indeed, continued improvement of such algorithms is important and a goal of many video surveillance researchers. It is an ultimately endless pursuit, however, and one not entirely necessary for video surveillance to deliver on a new and useful promise of providing direct, on-demand intelligence to end users. In the age of big data, it may be more important to employ advanced video analytics for the purposes of streamlining search and discovery or expediting end-user visualization [12] and supporting assisted (versus automated) exploitation.

Bernhard Rinner: One of the key challenges of current video surveillance is to transfer the technology from laboratory settings to real-world environments. What really counts here—from a customer's perspective—are the analysis capabilities, the overall performance and robustness, the ease of deployment, and finally the price. One could take a pessimistic view and argue that there has not yet been a killer application identified that is able to fulfill these high expectations. From an optimistic perspective one could acknowledge the tremendous progress achieved so far and envision a breakthrough by exploiting the advances in many related fields including sensors, fusion, and embedded computing, just to name a few. Assisted living, entertainment, or multimedia are novel application domains, which could benefit from video surveillance technology. Maybe we will see a killer application popping up in one of these related applications.

James Ferryman: Video surveillance has promised too much to many.

While some claims of what video surveillance can achieve are fiction, with advancements in sensors [13], processing hardware, and algorithms, there have been many successes in the design and deployment of video surveillance systems. For example, automated surveillance has been successfully deployed for traffic management (including detection of incidents), for automatic border control (e.g., iris recognition at U.K. borders), and for determining footfall (counting the number of people who enter a shop or business during a given time.) In the U.S. National Institute of Standards and Technology's (NIST's) assessment of biometric face recognition in still images, the error rate halves every two years. In 2010, the best face recognition method matched 92% of mug shots to one out of 1.6 million images [14]. More generally, automated surveillance can robustly detect "basic" events (e.g., a person entering a forbidden zone), global events (e.g., overcrowding), simple interactions (e.g., vandalism) based upon predefined semantics.

However, expectations from surveillance systems have only been met to a limited extent. A significant issue for users is that most deployed surveillance systems suffer from a too-high false alarm rate in detection of interesting events within the surveillance scene.

Taking critical infrastructure protection, a concrete example of current limitations of perimeter surveillance is provided by the US$1 billion SBInet "virtual fence" program for automatic monitoring of the warning at the U.S.–Mexican border [15]. This program was cancelled in 2011 because despite immense investment, the system was unable to distinguish between humans, cars, and animals. The false alarm rate due to animals and environmental factors was too high, and the system had poor performance in bad weather.

In counterterrorism, the London bombings of 7 July 2005 is an example of a major investigation undertaken by the U.K. Metropolitan Police Service, where CCTV footage was vital in understanding the sequence of events: 90,000 hard drives and videotapes from CCTV systems were seized totaling more than 6,000 hours of CCTV footage. Aside from the interoperability issues involved in collecting large volumes of CCTV, no use of video analytics was made (or were available for use) in the investigation due to the challenges involved in robustly recognizing the (same) individuals across large numbers of cameras and over time, among other needs. All of the CCTV footage that was deemed useful to the investigation was reviewed manually. It took four to five days to identify the bombers.

Apart from issues involved in minimizing false alarm rates and robustness in classification capability and under environmental variation, the single largest issues, which hamper surveillance, are the limited cognitive and adaptation skills of current systems. There is still no universal approach to the design of systems that exhibit "intelligence" in complex dynamic environments where a large number of events and activities can occur. As mentioned above, current surveillance systems have problems interpreting a scene as a human does with all the complex reasoning that this ensues. Tracking of individual people in crowds, keeping track of moving objects that are temporally occluded, and tracking and understanding interactions between multiple targets are further challenges. Illumination changes due to clouds, shadows, vibrations, or dirt on the sensor also cause problems. Interpretation of behavior is a higher level of cognitive skills that also is entirely missing in current systems. No visual sensor is capable of 24 hours a day, seven days a week, and 52 weeks a year operability under all these conditions.

Hence, many challenges remain for researchers and academics in the video surveillance community.

Brian C. Lovell: The major problem with current video analytics systems is their high false alarm rates and the difficult installation, configuration, and management. Many potential adopters of video analytics have experimented with these systems and then turn them off due to the false alarms.

Much research on video analytics has concentrated on improving security where there is no clear financial return. Management is unlikely to adopt new technologies simply to improve security. However if the video analytics systems can be used to replace a guard or to generate revenue, it is far easier to encourage adoption.

Péter L. Venetianer: For a variety of reasons, so far video surveillance has not met all the expectations.

Automated video surveillance applications initially focused primarily on outdoor perimeter protection, especially for critical infrastructure, such as power plants, airports, oil refineries, etc. Such perimeter protection scenarios frequently offer clear, unobstructed views with few moving targets, making the task appear tractable for intelligent video surveillance systems. Watching such surveillance videos for an extended period of time is a mind-numbingly boring task. Studies show that guards can effectively detect events of interest only for a very limited amount of time, making the problem an ideal candidate for automation. In reality, however, very few installations met user expectations, and customers lost trust in video surveillance. While it is easy to blame over-promising salespeople and Hollywood movies for setting unrealistic expectations, the problems are deeper, related to the limitations of today's systems. Even if a security guard is not always paying attention, serious security violations are relatively rare, so the lapse in attention may have no consequences. On the other hand, the intelligent video surveillance system not only misses detections, but will have false alarms as well, which are immediately noticed by the users. In addition, video surveillance systems will also have some missed detections during staged evaluations, which often seem incomprehensible to the evaluators, hurting the reputation of video surveillance. In addition to such performance problems, even the cost saving benefit of video surveillance is often questionable—video surveillance can reduce the number of people monitoring video feeds, but cannot replace the responders and in lot of facilities these two tasks are carried out by the same people, so eliminating monitoring doesn't necessarily reduce the required headcount. Typically, security-related investment represents a grudge buy, often without a return-on-investment (ROI) calculation. Finally, the security industry is extremely conservative and resistant to innovation, making entry a lot harder.

The problems described above left video surveillance with a tarnished reputation and with a very limited install base. It also meant that the field had to look for a new killer application. Fortunately, the low cost of cameras and computing power is helping it make a comeback in a variety of applications. A key feature is less demanding performance requirements, better tolerating both misses and false alarms. A typical example is business intelligence applications, measuring customer traffic, entries, exits, dwell time in front of displays, and time spent in a line. In such applications, accuracy is not as crucial as in security, limited miscounting is tolerated, and errors may even cancel each other out. In addition, indoor installation usually results in more controlled lighting, higher resolution on objects, and more limited variety of objects, all simplifying the computer vision challenge. Another example is finding stolen or delinquent vehicles by scanning the license plates of parked vehicles: the application is not sensitive to some missed detections, and can even tolerate limited false alarms.

Peter Tu: Unfortunately, the vast majority of today's pixels simply end up on mass storage devices only to be analyzed after a troubling event has taken place. This has led to the assertion that today's video analytics and septic tank industries are comparable in size. On the one hand, a number of almost intractable problems need to be solved. On the other, you only have to be able to dig a hole.

One could argue that for video analytics a "killer app" has never quite emerged due to the fact that business models for high returns on investment remain elusive. However, I feel that a greater concern to the community is that when confronted with real world complexity, many of our systems are simply not sufficiently robust and often fail in an ungraceful manner. This is a question of technical maturity, to which I would like to consider various aspects of how we, as a community, go about our research.

One could argue that one of our main approaches to driving progress is through the benchmarking of systems against publicly available data sets. It is almost impossible to publish a paper without comparing one's results with respect to such databases. I wholeheartedly applaud groups such as PETS, NIST (face grand challenge), and TRECVid for their tireless efforts in disseminating such data and providing forums for comparison of results. However, ROC curves are only one measure of success. I would argue that another measure of technical maturity is the question of whether or not one can rely on a given technology. The ability to perform on a given data set does not necessarily imply that the system can perform well when confronted with unforeseen imaging conditions and complexity. To this end, I would propose the instrumentation of a number of real-world sites such as airports, hospitals, city streets, parks, and shopping malls. Such sites would generate continuous streams of live video that could be accessed by any researcher in a real-time manner. Given access to such assets, I think that many researchers will be in a much better position to determine whether or not they can rely on their technology.

One of the key aspects of evolved systems is the ability to consolidate past progress. Unfortunately, this type of technological integration remains problematic for the video surveillance community. In many cases the ability to simply reproduce the results of another group remains problematic. To some degree this is an issue of complexity as well as intellectual property. However another issue maybe our unslacking thirst for novelty. The last thing that a new Ph.D. student wants to do is start working where a previous graduate student has left off. This is because at some point one simply hits the point of diminishing returns. So, one can either jump to a new area, where fresh results are easier to generate or one can simply hit the delete button and start from scratch. One solution might be to consider the construction of "super systems." Developers could insert their technologies into these entities resulting in hundreds or even thousands of trackers, face recognition engines, object classifiers, and behavior analysis systems, all working in tandem. What could emerge is some sort of oracle that would determine how best to exploit the individual advances of our collective community. By preserving and integrating in this way, we may unleash the forces of unforeseen symbiosis and the insights that can only be gleaned from such a comprehensive approach.

Anthony Hoogs: I strongly second Peter's call for more realistic, complete data sets for researchers to develop and test surveillance algorithms. As Peter and others indicated, algorithms that are successful in lab and research settings often fail badly under diverse real-world conditions. Why? Are the algorithms, or their developers, not smart enough? No. The answer is simpler, I think, and perhaps more disturbing: most algorithms are not designed to work under all real-world conditions. Researchers don't set out to limit their algorithms intentionally, but usually they do not try to make them work under adverse conditions. Instead, "difficult" conditions are avoided so that reported accuracies are high enough for publication. I cannot remember the last time I saw a video surveillance paper that showed results on a scene with rain, or snow, or blowing dust, or water on the lens, or horrible video quality from transmission dropouts, or image plane artifacts. Occasionally a paper will appear that tries to deal with one or more of these conditions independently, but not in the context of an end-to-end system.

Part of the problem is that we don't have data sets and a reward system for researchers to tackle these challenges. This is easily solved, and I encourage funding managers to prioritize the creation of highly diverse, real-world, research data sets exhibiting the full range of conditions. In the past 15 years, the computer vision community has vastly improved its scientific rigor by embracing common data sets and comparative evaluation. Introducing a comprehensive, freely available, unrestricted surveillance data set will significantly advance the state of the art. We created the VIRAT Video Data Set [16] with this in mind, but while it has a large diversity in scene content, it has the same shortcoming as other data sets in that it was not intended to sample the full range of imaging and scene conditions.

In terms of applications, I expect that easily monetized applications will drive most surveillance research in the next ten years. Commercial security is not in this category, as it is an overhead expense and difficult to calculate its ROI. Government security applications, particularly in military and intelligence, will likely continue to lead the funding and consequently drive the research in video for security. Retail applications and funding should continue to increase rapidly as capabilities mature, and clear benefits emerge in more domains.

Today, there are many domains where some level of automation in video analysis would save costs and improve operations, but it has not been adopted. This problem is not unique to video, and it will eventually resolve itself, although the pace of this can be frustrating. I have personally experienced two or three situations in the past ten years where semiautomated analysis capabilities were not adopted by their intended customers, despite clear evidence and user evaluations that indicated they would be highly beneficial. There are many mundane factors including funding limitations and politics, but there is also a fear of risking investment and careers on yet another attempt at automation that is probably oversold.

Technically, I'm a strong believer that we will continue to make rapid progress in the fundamentals underlying video analytics—tracking, object detection, and recognition including reidentification, video scene understanding, anomaly detection, and so on. If we focus our research in these areas on the problems of difficult scene and imaging conditions, we will develop more useful capabilities sooner.

Sharath Pankanti: Many challenges do exist as we continue our march to building our next generation surveillance systems, especially, relating multiple cameras, extracting details in crowded scenarios, assessing anomalous behavior, leveraging use of other sources of information including crowd-sourcing, social media, and improving collaborative environments among operators.

François Brémond: I totally agree with the previous answers. Video surveillance is a complex application field and a difficult business for many reasons. First, there is insufficient understanding of performance due to complex conditions of use, which are depending on many parameters. Besides, potential hard/middle/software constantly evolves. User needs and objectives are sometimes ill defined and changing. Video surveillance represents a segmented domain where many stakeholders such as system integrators, camera providers, public bodies, varieties of customers, network companies, insurance companies, and citizen associations interfere recurrently. Finally, the margin of profit is low yet it requires costly investments for mostly risk mitigations. There is a high competition between companies. To meet user expectances, this is an optimization problem between the video condition types, the technologies, the resources (processing power, network) and the user needs. Unfortunately, too few tools are available to improve this understanding.

Often a successful application has to find the right balance between structured scene (constant lighting, low people density, repetitive behaviors), simple technology (robust, low energy consumption, easy to set up, to deploy, to maintain, to extend), and strong motivation: fast payback investment, state regulation (e.g., tunnel, swimming pool), large market (user consumption), and affordable solution: US$150–$5,000 per smart camera.

Moderator: What do you think video surveillance systems will look like in the near future?

Bernhard Rinner: There has been tremendous progress achieved in automatic video analytics such as motion analysis, object detection and tracking, activity recognition, and identification. However, there is still a lot of research ahead until video surveillance is able to provide the same reasoning capabilities as the monitored persons are able to do. So, human visual cognition sets the bar—quite some challenge given the fact that we have been trained by evolution for thousands of generations. The expectations are high; achieving steady progress is crucial to avoid the situation a different technology, whose expectations were also strongly driven by human capabilities, has experienced some time ago: artificial intelligence (AI).

Sharath Pankanti: The existing systems are already attracting attention in mainstream media for their unprecedented functionalities [17], [18]. Given the emerging need and substantial commercial market, I am hopeful that surveillance will be hailed as the killer application not only for computer vision but also for AI.

Péter L. Venetianer: Ultimately, video surveillance will succeed in applications where it provides a clear benefit to its users; it is intuitive and easy to use, with robust performance. The biggest problem currently is that most systems are trying to achieve too much, and in environments with only few constraints performance is too poor and unreliable to be useful. Acceptance will likely start with some specific niche applications, with well-defined, predictable environmental constraints. Moderately priced special sensors, such as low-resolution thermal or stereo cameras that help overcome basic problems like shadows may also play a significant role.

James Ferryman: Surveillance systems of the future demand distributed, network infrastructures, greater robustness and adaptation in their vision algorithms and real-time 24 hours a day, seven days a week, 52 weeks a year operation delivering an enhanced level of sophistication in their semantic output while respecting social, legal, and ethical considerations.

Brian C. Lovell: Just as the Internet Protocol (IP) connected the world's computers 20 years ago, the advent of IP camera networks replacing the aging analog systems allows the interconnectivity of millions of surveillance cameras. This will lead to citywide surveillance networks, which will drive the need for video analytics such as robust noncooperative face recognition and person and vehicle tracking.

Due to the slow uptake of IP and the development of truly connected networks, I see both video surveillance and video analytics as being in their infancy. I anticipate that eventually every household will have surveillance video cameras and use easily configured reliable cloud-based video analytics to monitor activities in the home. These smart homes will monitor activities, recognize every householder, and provide services and security accordingly.

François Brémond: I believe more success stories would be possible with today's video analytics technology (especially if customers have direct access to engineers and not to commercials that oversell technologies) and there is still a high potential due to increasing resources (processors, sensors, smartphones, databases, cloud…) and increasing needs (uncontrolled violence).

[IF WE AS A COMMUNITY CAN FIND A WAY TO MORE EFFECTIVELY COORDINATE OUR EFFORTS, THEN THE DREAM OF VIDEO SURVEILLANCE MAY FINALLY BE WITHIN OUR GRASP!]

Shiloh L. Dockstader: Funding remains a major challenge for video surveillance research and development. Military and homeland security applications often involve the development of singular systems for dedicated use and possess a limited user base. Commercial entities often struggle to justify the ROI for video surveillance and security research, further slowing progress. To truly address this challenge, the financial motivation for investing in video surveillance research and development needs to increase, perhaps by expanding the size of the user base. This might be accomplished by better leveraging video processing and exploitation investments in tangential application areas like computer graphics, gaming, and entertainment.

Anthony Hoogs: With few exceptions, what is funded is what researchers work on. Whichever applications community continues or increases its investment in video automation will reap the benefits. I expect the U.S. military and intelligence communities will continue their substantial investment in video analysis, as indicated by the Undersecretary for Defense, Intelligence at the 2012 GEOINT Conference when he stated that his top research priority would continue to be unmanned aerial vehicle video systems. Increased coordination between the various U.S. military and intelligence agencies, through a more open, common, broad-based research procurement approach, would massively increase their ROI.

With the ubiquity of consumer video, I expect commercial funding will increase in collateral areas that will incidentally benefit video surveillance. Automated cataloging of consumer video will be a huge driver, as it is for consumer photos with products such as Picasa.

Technically, the confluence of computer vision, machine learning and AI should yield substantial improvements in video analytics. Understanding human situations, intentions and actions is not sufficiently pursued or funded today, in favor of advancing the more traditional areas of tracking and recognition. While those areas are important, adjusting the funding balance towards what to do with tracks and objects would yield more progress sooner.

Peter Tu: It could be argued that the challenges facing video analytics is comparable to that of hard physics. While a single particle physics team may have hundreds or even thousands of members, our community remains a cottage industry. If we as a community can find a way to more effectively coordinate our efforts, then the dream of video surveillance may finally be within our grasp!

MODERATOR

Fatih Porikli (fatihporikli@ieee.org) is a Distinguished Research Scientist at

Mitsubishi Electric Research Labs (MERL). He received his Ph.D. degree from the Polytechnic Institute of New York University. His work covers areas including computer vision, machine learning, video surveillance, multimedia processing, structured- and manifold-based pattern recognition, sparse reconstruction, biomedical vision, radar signal processing, and online learning. He has over 100 publications and 60 patents. He mentored more than 40 Ph.D. students. He received the 2006 R&D100 Award in the "Scientist of the Year" category (a select group of winners) in addition to three IEEE Best Paper Awards and five professional prizes. He was an associate editor for many IEEE, Springer, and SIAM journals. He was the general chair of the 2010 IEEE Advanced Video and Signal-Based Surveillance Conference and program chair of many IEEE events.

PANELISTS

François Brémond (francois.bremond@inria.fr) is a research director at INRIA Sophia Antipolis. He is currently leading the STARS team and previously he was the head of the PULSAR team. He obtained his M.S. degree in 1992 from ENS Lyon. He has conducted research work in video understanding since 1993 both at Sophia Antipolis and at the University of Southern California. He designs and develops generic systems for dynamic scene interpretation. He authored of more than 100 scientific papers published in international journals or conferences in video understanding. He has (co)supervised 12 Ph.D. theses and is a European Commission Information Society and French National Research Agency expert for reviewing projects.

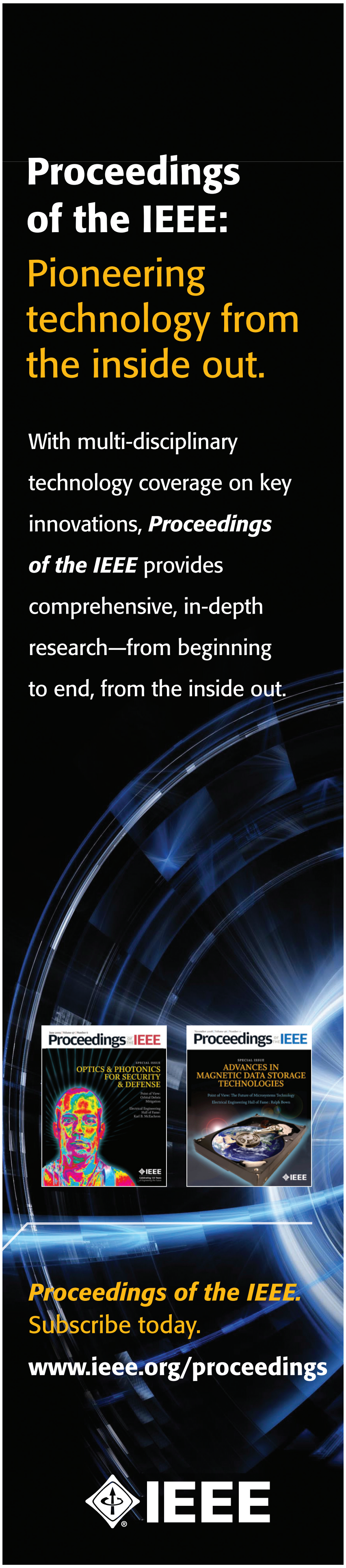
Shiloh L. Dockstader (shiloh.dockstader@exelisinc.com) is currently employed with ITT Exelis Geospatial Systems as the chief scientist for Geospatial Intelligence Services, where he leads research and development activities in the areas of motion imagery analysis, advanced processing and exploitation, and multisource information fusion. He provides engineering oversight for activities with and serves as a primary technical liaison to external organizations for ITT Exelis. He provides scientific advisory support to the National Geospatial-Intelligence Agency's (NGA's) InnoVision and Analysis Directorates, offering technical guidance to a number of NGA programs, divisions, and National System for Geospatial Intelligence research and development portfolio managers. He has experience architecting and operationally deploying advanced signal-based processing and exploitation systems for intelligence, surveillance, and reconnaissance applications.

James Ferryman (j.m.ferryman@reading.ac.uk) leads both the Computational Vision Group (CVG) and the wider Computing Research Group in the School of Systems Engineering, University of Reading, United Kingdom. CVG has a current focus on automated surveillance including CCTV analysis for safety, security, and threat assessment. He has received extensive funding from research councils in the United Kingdom, the European Union (EU), and industry and has acted as principal investigator on projects including UK EPSRC REASON (EP/C533402) examining robust methods for monitoring and understanding people in public spaces, and EU Framework 6 and 7 projects including SAFEE (onboard aircraft threat detection), ARENA (protection of critical mobile assets), and EFFISEC (efficient integrated security checkpoints). He has acted previously as the director of both the British Machine Vision Association and Security Information Technology Consortium, both of which he is a member.

Anthony Hoogs (anthony.hoogs@kitware.com) received his Ph.D. degree in computer and information science from the University of Pennsylvania in 1998. He founded and directs the computer vision group at Kitware, Inc., which currently has more than 30 members including 13 Ph.D students. Over the past 20 years, he has supervised and performed research in various areas of computer vision including: event, activity and behavior recognition; motion pattern learning and anomaly detection; tracking; and content-based retrieval. He has led numerous research projects sponsored by the Defense Advanced Research Projects Agency (DARPA), AFRL, ONR, I-ARPA, NGA and corporate funds. Recently, he was the overall principal investigator on two DARPA video analysis projects, VIRAT and PerSEAS, which included collaborations with multiple universities and research institutions. He has published more than 60 papers in computer vision, pattern recognition, artificial intelligence, and remote sensing.

Brian C. Lovell (lovell@itee.uq.edu.au) received the B.E. degree in electrical engineering in 1982, the B.Sc. degree in computer science in 1983, and the Ph.D. degree in signal processing in 1991, all from the University of Queensland (UQ). He is the director of the Security and Surveillance Group in the School of Information Technology and Electrical Engineering, UQ, and project leader of the Advanced Surveillance Project. He was the president of the International Association for Pattern Recognition (IAPR) from 2008 to 2010, Fellow of the IAPR, Senior Member of the IEEE, and is a voting member for Australia on the Governing Board of the IAPR. He is the general cochair of ICIP 2013 to be held in Melbourne, Australia.

Sharath Pankanti (sharat@us.ibm.com) is a research staff member in the Software Research Department at the Thomas J. Watson Research Center. He received a Ph.D. degree in computer science from the Michigan State University. He is manager of Exploratory Computer Vision Group at the Thomas J. Watson Research Center where he has led a number of safety-, productivity-, and security-focused projects. He is a coauthor of more than 80 inventions and more than 125 technical papers. He coedited the first comprehensive book on biometrics, *Biometrics: Personal Identification* (Kluwer, 1999) and coauthored *A Guide to Biometrics* (Springer 2004) which is being used in many undergraduate and graduate biometrics curricula. He is a member of ACM and a Fellow of the IEEE.



Bernhard Rinner (b.rinner@computer.org) is a full professor at the Alpen-Adria-Universität Klagenfurt, Austria, and deputy head of the Institute of Networked and Embedded Systems. Before joining Klagenfurt, he was with Graz University of Technology and held research positions at the Department of Computer Sciences at the University of Texas at Austin in 1995 and 1998 and 1999. His current research interests include sensor networks, embedded video and computer vision, pervasive computing, and UAV systems. He was a cofounder and general chair of the ACM/IEEE International Conference on Distributed Smart Cameras. Together with partners from four European universities, he has jointly initiated the Erasmus Mundus Joint Doctorate Program on Interactive and Cognitive Environments.

Peter Tu (tu@research.ge.com) received his B.S. degree in 1990 in systems design engineering from the University of Waterloo, Canada. He then earned his doctorate in 1995 from Oxford University's Department of Engineering Science. In 1997, he joined General Electric's Computer Vision Group, where he now serves as its principal scientist. He is currently focused on developing video analytic technologies including object detection, object tracking, scene understanding, behavior recognition, affective analysis and biometric-at-a-distance capture systems. He has been a principal investigator for the Federal Bureau of Investigation, National Institute of Justice, DARPA, and the Department of Homeland Security. He has 21 issued patents and over 50 peer-reviewed publications.

Péter L. Venetianer (pvenetianer@ObjectVideo.com) received the M.S degree in computer science from the Technical University of Budapest in 1992 and the Ph.D. degree from the Computer and Automation Institute of the Hungarian Academy of Sciences in 1996. He is the senior director of commercial science development at ObjectVideo. His responsibilities include managing the teams that design, develop, and test new technologies for the advanced video analytics product line of ObjectVideo. He performed research and development in automated video processing, including background modeling, segmentation, event detection and recognition, and activity inferencing. Prior to joining ObjectVideo in 2000, he worked on iris recognition at Sensar, Inc.; and he performed research on image processing, computer and machine vision, visual communication, and coordination between multiple agents at the University of Pennsylvania.

REFERENCES

[1] *Popular Sci.*, Sept. 1939.

[2] U.S. Patent 3114797, 1963.

[3] U.S. Patent 3590151, 1966.

[4] R. Hertzberg, "A cure for crime in the parks," *Sci. Mech.*, Jan. 1965.

[5] G. Smith, "Behind the screens: Examining constructions of deviance and informal practices among CCTV control room operators," *Surveill. Soc.*, vol. 2, no. 2–3, 2002.

[6] Special issue on Video Communications, Processing, and Understanding for Third Generation Surveillance Systems, *Proc. IEEE*, vol. 89, no. 10, pp. 1355–1367, Oct. 2001.

[7] [Online]. Available: http://www.news.com.au/national/one-of-six-people-seen-on-crucial-cctv-comes-forwardas-police-probe-abduction-theory-on-missing-jill-meagher-and-plea-for-witnesses/story-fncynjr2-1226482146903

[8] [Online]. Available: http://facewatch.co.uk/cms/

[9] (2011, Mar. 3). Only 1.85 million cameras in UK, claims ACPO lead on CCTV. [Online]. Available: http://www.securitynewsdesk.com

[10] R. Feris, B. Siddiquie, Y. Zhai, J. Petterson, L. Brown, and S. Pankanti, "Attribute-based vehicle search in crowded surveillance videos," in *Proc. ICMR*, 2011.

[11] S. Dockstader, M. Berg, and M. Tekalp, "Stochastic kinematic modeling and feature extraction for gait analysis," *IEEE Trans. Image Processing*, vol. 12, no. 8, 2003.

[12] Y. Pritch, A. Rav-Acha, and S. Peleg, "Nonchronological video synopsis and indexing," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 30, no. 11, pp. 1971–1984, 2008.

[13] High-def CCTV cameras risk backlash, warns UK watchdog. [Online]. Available: http://www.bbc.co.uk/news/technology-19812385

[14] (2012, Apr. 28). Video surveillance: I spy, with my big eye. *The Economist* [Online]. Available: www.economist.com/node/21553408

[15] U.S.–Mexico border: Efforts to build a virtual wall. [Online]. Available: http://www.bbc.co.uk/news/technology-19409682

[16] S. Oh, A. Hoogs, A. Perera, N. Cuntoor, C.-C. Chen, J. T. Lee, S. Mukherjee, J. K. Aggarwal, H. Lee, L. Davis, E. Swears, X. Wang, Q. Ji, K. Reddy, M. Shah, C. Vondrick, H. Pirsiavash, D. Ramanan, J. Yuen, A. Torralba, B. Song, A. Fong, A. Roy-Chowdhury, and M. Desai, "A large-scale benchmark dataset for event recognition in surveillance video," in *Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR)*, 2011.

[17] [Online]. Available: http://abclocal.go.com/wls/story?section=news/national_world&id=6138580

[18] [Online]. Available: http://abclocal.go.com/wls/story?section=news/special_segments&id=7294108

[SP]