Survey and Analysis of Multimodal Sensor Planning and Integration for Wide Area Surveillance

BESMA R. ABIDI, NASH R. ARAGAM, YI YAO, and MONGI A. ABIDI

The University of Tennessee

Although sensor planning in computer vision has been a subject of research for over two decades, a vast majority of the research seems to concentrate on two particular applications in a rather limited context of laboratory and industrial workbenches, namely 3D object reconstruction and robotic arm manipulation. Recently, increasing interest is engaged in research to come up with solutions that provide wide-area autonomous surveillance systems for object characterization and situation awareness, which involves portable, wireless, and/or Internet connected radar, digital video, and/or infrared sensors. The prominent research problems associated with multisensor integration for wide-area surveillance are modality selection, sensor planning, data fusion, and data exchange (communication) among multiple sensors. Thus, the requirements and constraints to be addressed include far-field view, wide coverage, high resolution, cooperative sensors, adaptive sensing modalities, dynamic objects, and uncontrolled environments. This article summarizes a new survey and analysis conducted in light of these challenging requirements and constraints. It involves techniques and strategies from work done in the areas of sensor fusion, sensor networks, smart sensing, Geographic Information Systems (GIS), photogrammetry, and other intelligent systems where finding optimal solutions to the placement and deployment of multimodal sensors covering a wide area is important. While techniques covered in this survey are applicable to many wide-area environments such as traffic monitoring, airport terminal surveillance, parking lot surveillance, etc., our examples will be drawn mainly from such applications as harbor security and long-range face recognition.

Categories and Subject Descriptors: A.1 [**General Literature**]: Introductory and Survey; I.2.10 [**Artificial Intelligence**]: Vision and Scene Understanding—*Motion, video analysis*; I.4.8 [**Image Processing and Computer Vision**]: Scene Analysis—*Motion, object recognition, range data, sensor fusion, tracking*; I.5.4 [**Pattern Recognition**]: Applications—*Computer vision*

General Terms: Algorithms, Design, Performance, Security

Additional Key Words and Phrases: Sensor planning, ubiquitous surveillance, cooperative sensing, wireless sensor network, multimodal sensing, threat object recognition, persistent surveillance

ACM Reference Format:

Abidi, B. R., Aragam, N. R., Yao, Y., and Abidi, M. A. 2008. Survey and analysis of multimodal sensor planning and integration for wide area surveillance. ACM Comput. Surv. 41, 1, Article 7 (December 2008), 36 pages DOI = 10.1145/1456650.1456657 http://doi.acm.org/10.1145/1456650.1456657

7

This research was supported by the DOE University Research Program in Robotics under grant DOE-DE-FG52-2004NA25589 and ONR N00014-03-1-0331.

Authors' addresses: B. R. Abidi, Department of Electrical Engineering and Computer Science, The University of Tennessee, Knoxville, TN 37996; email: besma@utk.edu; N. R. Aragam, School of Public Health, East Tennessee State University, Box 70267, Johnson City, TN 37614-1700; Y. Yao, M. A. Abidi, Department of Electrical Engineering and Computer Science, The University of Tennessee, Knoxville, TN 37996.

1. INTRODUCTION

The broad area of computer vision deals with the general problem of using and placing a sensor as the eye of a vision system to enable such tasks as automatic object characterization, manipulation, reconstruction, etc. The field of 3D object reconstruction from multiple laser range and visual images has received the most attention, as seen from a search of literature in sensor placement techniques by Price [2007]. A classic paper by Tarabanis et al. [1995] gives an excellent summary of literature of sensor placement strategies for computer vision in a limited near-field context of robotic inspection until 1995. More recently, Roy et al. [2004] discussed sensor planning in the context of 3D object recognition and scene analysis applications. In addition, Chen and Li [2004] also addressed the placement of active sensors in the context of robot vision. In this article, we are more interested in active sensing for object characterization and situational awareness in the context of wide areas.

Hu et al. [2003] recently surveyed the state-of-the-art in visual surveillance with dynamic scenes to understand and characterize object motion and behavior. They summarize and identify low-, medium-, and high-level visual surveillance. However, they do not seem to look into such central problems as long-range sensing requirements, modality selection, positioning, fusion, and collaboration of sensors needed for a wide-area surveillance application. In this article we hope to address this need and survey the state-of-the-art in wide-area surveillance with focus on multisensor placement and integration to provide an in-depth analysis and available strategies in literature. The main motivation for undertaking this research can be summarized as the need to monitor wide areas such as harbors, airports, and large public parking facilities. Adaptive sensor placement and collaboration in a dynamically changing, wide 3D scene combined with such real environmental features as night/day, rain/wind, fog/haze/clear conditions, etc., is not only a challenging research problem to solve, but is also of extreme importance when it comes to threat assessment and situational awareness. Figure 1 illustrates a large harbor channel, where ships and barges enter and leave, as an example of a wide area in need of surveillance. We need to understand the factors involved in long-range and covert collection of high-resolution information on the incoming ships and barges with suspicious cargo (e.g., explosives triggered by remotely controlled switches) into a harbor channel for situational awareness and potential threat assessment. While a number of independent studies and research publications have been commissioned by the U.S. Government after the tragedies of September 11, 2001, a large number of analytical problems in sensing for security and surveillance remain unsolved. For example, what is the highest resolution at which dynamic objects in motion can be detected for potential threat from a given distance? How could sensors be placed in a known network to cooperatively detect an impending disastrous situation? And so on.

The scope of this survey is necessarily broader than it has been for survey papers of the past dealing with scenes covering a limited area and the placement required for industrial robots in machine vision applications. It should be mentioned that the topic of this research is to survey the subject matter widely and determine possible strategies that may exist in formulating a path towards solving the general sensor placement and data integration problem in a wide-area surveillance application. Therefore, in addition to computer vision literature, we have used the literature available from GIS, remote sensing, surveying and terrain planning/guarding, and operational research. Therein, the sensor placement problem appears as a general optimization problem involving a determination of a largest surface coverage by a single or multiple viewpoints on a 3D terrain. Typically, such an approach has been used to solve for an optimal placement of watch towers and radio transmission towers. It should be noted that while the general scope extends to outdoor areas on 3D terrains for the determination of heights

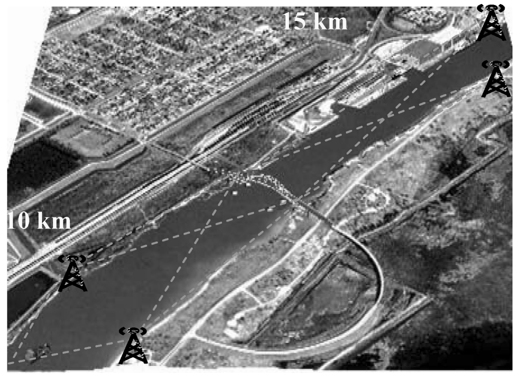


Fig. 1. Long-range sensing of wide-area surveillance for a case study of the harbor and bridge area in Port Arthur, Texas.

at which potential sensors may be located, the actual orientation of the sensor itself will use techniques and algorithms from machine vision and other computer graphics applications, where occlusions and illumination problems pose significant challenges to viewing an object of interest.

The main contributions of this article are listed as follows: (1) a survey and summary of sensor selection and attribute requirements for wide-area surveillance; (2) identification and analysis of visibility and coverage for sensor positioning; (3) review and discussion of sensor data fusion and cooperative intelligent sensing for situational awareness and threat assessment; and (4) synthesis and recommendations for intelligent and optimal sensor placement and deployment in wide areas for autonomous surveillance.

The remainder of this article is organized as follows. Section 2 describes the wide-area and long-range sensing requirements in terms of sensing modalities, sensor attributes, and parameters. Section 3 delves into sensor coverage and visibility analysis from multiple sources, namely, use of techniques from GIS, remote sensing, and cellular communications, in addition to those from computer vision. Section 4 discusses sensor collaboration and data fusion. Finally, Section 5 synthesizes the available strategies, analyzes the unique requirements, and offers sensor planning recommendations. Section 6 presents the conclusions for long-range surveillance and wide-area monitoring.

2. REQUIRED ATTIBUTES FOR SENSORS IN WIDE-AREA SURVEILLANCE

Before delving into specific requirements and constraints involved in the sensor placement for wide-area surveillance and monitoring, a general problem definition will be given. The survey paper by Tarabanis et al. [1995] considers the problem in the context of robotic inspection as threefold: (1) determining the number and type of sensor parameters, (2) figuring out the sensor and object models, and (3) determining the constraints to be posed on the detectability of a feature. However, the sensor modality for wide-area surveillance cannot be restricted to visual or optical sensors only. In this section, the available and necessary sensor modalities beyond just the visual are described along with their underlying parameters, attributes, and modes of operations. We will focus on four most applicable modalities of sensing: (1) radar range, (2) visual (optical), (3) thermal (infrared), and (4) laser range. Since wide-area surveillance generally involves long-range sensing in an uncontrolled outdoor environment, at ranges of 10 to 15 kilometers, the object visibility by the sensor is of utmost importance and becomes the major attribute under consideration.

In general, long-range sensors may be classified into active and passive types. Active types like radar range sensors use an active signal (e.g., radio waves) to bounce off a target and the reflected waves are analyzed to figure out the nature and dynamics of an object under surveillance. On the other hand, passive types like an infrared sensor depend on the detection of naturally emitted and reflected radiation from an object under surveillance. While both radar and infrared sensors provide long-range observation capabilities, active sensing via radar has distinct advantages like operation in bad weather (rain, fog, haze, and smoke) conditions and is not susceptible to the jamming conditions that are inherently possible in the case of thermal (infrared) sensors. Nevertheless, active sensing using radar does have its own disadvantages in terms of expensive noise and clutter elimination problems and the higher system cost for the radar hardware and software.

2.1. Radar Range Sensors

As the name itself implies, radar (Radio Detection and Ranging) can detect as well as range a target from a distance. There are several types of radars according to their intended applications. Some of them are available for wide-area surveillance. Depending on the cost, we could also find man-portable and other airborne radars suitable for wide-area surveillance. Table I shows a classification of the major types of radars along with their application, technology, resolutions, and ranges.

For wide-area surveillance applications, both Synthetic Aperture (SAR) and Moving Target Indicator (MTI) types would be useful in detecting the object of interest. In addition, man-portable radars and their technology point to the state-of-the-art in military and civilian surveillance systems. Some researchers [Kuttikkad et al. 1997] claim that SAR is becoming a sensor of choice for remote sensing and wide-area surveillance. SAR usage, along with some other electro-optical (e.g., infrared) sensors, is also being pursued by some military contractors in their turn-key systems [DRS 2008]. Schooley and Thompson [2004] have described an Ethernet-based integrated surveillance system for homeland security and homeland defense using radar and electro-optical sensors.

Essentially, radar range sensors provide range and extent information about the object of interest and may be used to detect and track an object from a long distance. The idea in using them in wide-area surveillance is to have a ground truth map of the area to be monitored and to set up range markers or rings around a known central point. By overlaying the range map with an electro-optical image (infrared or color visual) and calibrating using rigorous wide-area (or site) models, we can set up security zones accurate to within a meter and label the image clearly for monitoring the activity, such as vehicle and people traffic at any given time or round the clock. From the a priori knowledge of objects any security breach may be detected by tracking the objects using a long-range SAR sensor which places the targets on the pre-established area map with security zone markings. The higher resolution sensors such as high zoom PTZ cameras can then be called upon for detailed analysis of the threat.

Researchers at Sandia National Labs have recently developed prototype systems that enable definition of 3D volumetric security zones, and an advanced exterior multimodal sensor (AES) [Ashby and Pritchard 2004] that integrates thermal infrared, visible, and microwave radar sensing technologies. The 3D Video Motion Detection (3DVMD) system [Nelson 2004] provides switch closures for an interface to an alarm communications and display system, as well as information on the size, location, and direction of movement of moving objects. The AES, on the other hand, fuses sensor data from three sensors while scanning the wide area in 360° every second to collect the surveillance data. Both of these technologies are now being transferred to commercial manufacturers for production of low-cost, reliable, and high-performance wide-area surveillance sensors.

Table I. Classification of Radar Sensors

Type	HF Surface/Skywave	SAR (Synthetic Aperture Radar), ISAR	mm Wave Radar	MTI (Moving Target Indicator)	Man-portable
Application	Detection and tracking	Imaging and detection of stationary targets	Detection, imaging, tracking for collision avoidance	Tracking of moving targets	Target acquisition and tracking
Target	Boats, ships, low flying aircrafts	Scenery landscape	Aircraft and ground vehicle	Any moving target	Long range (up to 20 Km) moving targets
Waveform	Pulsed Doppler, FM /CW pulses	Linear FM pulses	Pulsed Doppler	CW	CW, pulsed and hybrid
Frequency	3–30 MHz	10 GHz	95 GHz	1.8–2.8 GHz	8–18 GHz
Resolution	Variable (∼3m)	0.5 m	1.5 m	Variable (5m)	Approx. ∼50m
Range limit	Shore-line −500, 500–3.5 Km	2–17 Km[a]	50–3000 m	Approx. ∼20 Km	Approx. ∼25 Km
Example	Raytheon Canada SWR-503	Sandia Twin Otter SAR	Commercial	Any radar	Any MSTAR

[a]The listed range is for airborne applications. The sensor also provides spaceborne-oriented scenarios with the range limit extended to 785 Kilometers. Since this survey focuses on wide-area surveillance, the spaceborne applications are beyond our scope.

Radar range sensors have also been employed by the U.S. Navy, the U.S. Customs, and the U.S. Coast Guard for wide-area surveillance of drug traffickers on speed boats and low-flying aircrafts. Two such applications were reported recently in literature [Ponsford 2004; Seastrand 2004].

2.2. Thermal (Infrared) Sensors

With optical sensors, any surveillance activity would be limited to well-lit conditions only, whether the lighting is provided by the Sun during the day or by artificial and planned lighting of dark areas during the night and/or day. On the other hand, thermal (or infrared) imaging uses the heat radiation or the infrared spectrum that is independent of the ambient light in the area to be imaged, and therefore useful for night-time and hidden-area surveillance. Infrared imaging technology has been routinely used for night vision by the military and civilian systems for surveillance and has now become a mature main-stream technology.

Both astronomy and medicine have employed thermal imaging in a wide array of applications and the literature points to a large amount of varied types of thermal sensors in use today. For long-range imaging, the military and the remote sensing community have developed systems to image a human at a distance of 8 kilometers or more!

Thermal sensors rely on the emission of infrared energy in the 1 micron to 100 microns wavelength range. Some of the important parameters that specify a thermal sensor are measurement temperature range, spectral range, accuracy, resolution, and maximum frame rate. The measurement temperature range is the range of temperatures to be measured on the object, and the cooler the object, the more bulky and expensive the sensor because of the needed cooling of the detector itself using liquid nitrogen, etc. Three commonly used detector types are flying spot, scanning, and line array. These correspond to the number of detectors[1] one for flying spot, a single line of detectors in scanning type, and a 2D array of detectors for the array type. Thermal sensors have both a horizontal and vertical spatial resolution that represents the number of pixels in either direction. The greater the resolution, the smaller the unit of temperature that can be measured.

Most of the literature available on using infrared sensors in wide-area applications seems to be concentrated in the defense and military reconnaissance domains. References [Crebolder et al. 2003; Office of Technology Assessment 2003] represent a couple of white papers describing the general role of infrared sensors in large military surveillance systems. Kaplan and Scanlon [2001] summarized the state-of-the-art in thermal imaging in a comprehensive report. Their findings are summarized in Table II. In surveillance and object recognition applications, thermal sensors have been employed to fuse thermal data with optical data in order to detect and classify features on a human face [Heo et al. 2004; Kong et al. 2005], etc.

2.3. Visual (Optical) Sensors

Currently, CCD-based digital cameras are by far the most prevalent mode for imaging using optical sensors. Image resolutions vary widely, starting from low NTSC standard resolution of about 460×350 and going up to 21 Megapixels (roughly 5120×4096) available in some high-end cameras. The technologies used in the sensor are usually CCD or CMOS and a wide variety of both optical and digital zooming capabilities are built in to enhance the range and applicability of the cameras. The maximum optical

[1]For example, Sierra-Pacific Systems (www.×20.org).

Table II. Performance and Suggested Applications of Three Categories of Modern IR Thermal Imagers

Characteristics	Uncooled Detector	Cooled MWIR	Cooled QWIP
Spectral range	Flat – filtered to 7 – 15	Selectable, 2.0 – 5.5 μm	Selectable, 7 – 9 μm
Typical frame rate	50 / 60 Hz	50 / 60 Hz	Selectable, 50/60 to 900 Hz
Typical thermal sensitivity	80 milliKelvins @300K	70 milliKelvins @300K	20 milliKelvins @300K
Typical applications	• General purpose	• Spectrally selective sensing	• Automation
	• Condition monitoring	• Process monitoring (plastics, glass, furnace, etc.)	• Military research
	• Predictive maintenance	• Condition monitoring	• Product development
	• Building & structural diagnostics	• Predictive maintenance	• Biomedical studies
	• Process monitoring	• Building & structural diagnostics	• Non-destructive testing (high speed, high sensitivity)
	• Security, surveillance, law enforcement, fire fighting	• Non-destructive testing (moderate speed, moderate sensitivity)	• Aerospace research
	• Biomedical measurements		• High resolution thermal mapping
	• Non-destructive testing (moderate speed, moderate sensitivity)		

MWIR: medium wavelength infrared and QWIP: Quantum well infrared photo detector

zoom capability to date provided by commercial PTZ cameras is 35×. IDC, a market research firm, has done a general market survey of digital cameras and scanner usage by consumers [IDC 2006]. Similar studies and surveys for surveillance video cameras are also available in literature.

IP cameras have attracted increasing attention in recent years mainly because of their easy installation and user-friendly interface. IP cameras make use of existing infrastructure built for the Internet and communicate with the control center based on standard Internet protocols. One major disadvantage of using IP cameras lies in the limited bit rate. Images with higher resolution frequently result in a frame rate less than 30 frames per second. With the development of the Giga-Ethernet connection, the transmission rate can reach 1Gbps, which is capable of handling streaming image data and of providing reliable transmission.

Two major parameters are associated with vision sensors: the observation viewpoint and the illumination point. The observation viewpoint actually consists of the position and orientation of the observer and the optical parameters associated with the sensor itself, such as the lens aperture. Thus, in general, two types of sensor parameters may be defined, namely geometric and optical. The geometric parameters in turn consist of positional (x, y, z) and orientational (pan, tilt, and swing angle) parameters. The optical parameters that would be important to know are the focal length of the lens, the aperture size, and the principal point.

Long-range imaging or sensing in general brings forth the issue of object resolution and size when viewed from a very long distance. In order to properly detect a given feature of an object from a long distance would require external mechanisms such as zoom lenses attached to the sensing equipment. Two of the aspects of zooming are focus and resolution of the zoomed-in image and their interplay in determining the clarity with which an object feature can be imaged for to be analyzed for an impending threat situation. Furthermore, zooming on demand would be required in wide-area surveillance and this brings up the question of multiresolution imaging capability of the sensor and its inherent control. Additionally, such optical parameters as exposure time, gain of the video signal by a camera amplifier, etc., will also become important in sensor planning.

The illumination parameters are also of two types: geometric and radiometric. The geometric parameters are the position, orientation, and characteristics of the illuminating beam (e.g., conical volume of the beam). The radiometric parameters are such factors contributing toward radiant intensity or energy, spectral distribution of intensity, and any polarization parameters involved in the control of the beam. In general, for outdoor wide-area applications, illumination parameters do not come into play and are unimportant. In summary then, just the sensor parameter problem becomes an n-dimensional problem to solve mathematically, assuming that we can develop a suitable sensor model that can succinctly describe all the aforementioned parameters.

In wide-area surveillance, the use of multiple visual sensors to track an object of interest is common. In particular, fisheye (or omnidirectional) lenses are employed to glean a global 90/360° view of a wide area under surveillance. In general, the fisheye sensor is used as a master along with one or more PTZ sensors as slaves to track an object [Cui et al. 1998; Scotti et al. 2005; Yao et al. 2006]. The idea here is that when a particular PTZ camera tracking an object is suddenly blinded by glare or an obstruction in the path of the moving object, it can handover control to the fisheye sensor to continue tracking until another PTZ camera is able to take over. Alternatively, the fisheye provides a low-resolution image and is good at pointing the general location, whereas the PTZ with high zoom factor would get us the feature we are interested in observing. The assumption here is that all the participating sensors are precisely calibrated to be able to localize an object to be tracked in the known confines of the wide-area geometry.

Additional literature on surveillance systems using omnidirectional cameras and videos may be found in Nayar and Boult [1998], where the development of omnidirectional sensors and real-time algorithms for tracking and depth estimation for distributed surveillance is reported. Power and Boult [2000] described an experiment to measure the effectiveness of different interfaces (team versus single-operator interface) in target identification and localization tasks. Babaguchi et al. [2002] presented a media system called OVISS which visualizes and summarizes omnidirectional surveillance video to a viewer. In Chen and Yang [2002], the authors proposed an approach for monitoring human activities in an indoor environment using an omnidirectional camera. Finally, Vanijja and Horiguchi [2001] developed an optimal method to extend the degree of freedom of a user's movements in a virtual environment (used in remote monitoring) of the scene imaged by an omnidirectional camera.

2.4. Laser Range Sensors

In addition to radar, thermal, and optical technologies, laser range sensor technology that measures depth to a very high degree of accuracy (in terms of a few centimeters and millimeters) is available for various-range imaging. Typically, a fine laser is used to sense an object from a distance by using the time required for the signal to bounce back from the object and arrive at the source. The simplest type of a range sensor may be found in typical handheld police automobile speed detection guns. More sophisticated laser rangers are used routinely in 3D reconstruction of scenes and objects in image processing, computer graphics, and computer aided mechanical/architectural design.

Table III. Medium-Range Laser Scanners

Producer	Model	Range (meters)	Principle
Riegl laser measurement systems	LMS-Z210	450	Time-of-flight (tof)
Calidus precision systems	Callidus	80	tof
Metric vision	MV 200	60	tof
Zoller+Froehlich GmbH	LARA	55	tof
Cyra Technologies	Cyrax 2500	50	tof
Mensi	GS200	25	tof
3rd Tech	DeltaSphere	12	tof

For surveillance applications, the military has employed range sensing in a number of applications to-date. The important parameters in laser range sensing are laser beam characteristics such as size, divergence and distribution of light, and the wavelength of the laser. Most laser range sensors actually have an imaging camera in addition to the laser itself, which is basically used to acquire a color image for texturing the object under surveillance or observation. A summary for medium range scanners applicable to wide-area surveillance is presented in Table III.

For ranges in excess of a kilometer that we find in wide-area surveillance applications, the literature points to a number of military systems using a laser range finder in concert with an infrared sensor and an electro-optical sensor. The laser is not used for scanning the object under surveillance in these systems. However, Dayton et al. [2000] demonstrated the use of a LADAR (Laser Radar for Laser Illuminated Imaging) for nighttime surveillance at distances in excess of 20 Kilometers. A number of groups have investigated the use of LADARs in tactical systems [Miller et al. 1998; Monson et al. 1999; Steinvall et al. 1999; von der Fecht and Rothe 1998].

Dayton et al. [2000] pointed out that a number of factors limit the maximum useful range of a LADAR for high-resolution imaging. These include atmospheric turbulence, coherent speckle effects, aerosol back-scatter, and the laser energy required to produce a return photon signal of a particular level to be recognized. Another interesting project using laser illumination can be found in Greer et al. [1997].

3. SENSOR PLANNING AND POSITIONTION FOR WIDE-AREA SURVEILLANCE

Sensor placement and planning in computer vision, computer graphics, and visualization has been researched for a relatively short time (since the mid-1980's). In this section, we will address the following issues: (1) coverage analysis, (2) feature detectability constraints, and (3) sensor positioning in wide areas.

3.1. Coverage Analysis

In this section, we examine the problem of wide-area surveillance from the viewpoint of coverage. In particular we look into several strategies available from areas outside traditional image processing and computer vision. In addition, we discuss coverage with multiple sensors that could be either fixed or in motion relative to the object of interest. There are alternative areas such as photogrammetry and remote sensing, geography, and operational research where the visibility problem in a wide area has been studied. In so far as the placement of a sensor is concerned, this problem, especially in wide-area, applications, is the same whether the sensor is trying to capture an image or a signal from a point in a wide area. One example of this would be the placement of a radio or cell phone transmitter tower in an urban or rural area. Another would be selecting an optimal position to place a watch tower in a forest fire prevention system.

The common problem solved in these alternative techniques is the one of finding the largest coverage area from a given observation point in 3D space. Usually, we determine the optimal height at which the sensor needs to be placed so that a viewing cone subtended from that point in space will encompass the largest elliptical area on the ground. Depending on the regularities of the terrain, a point on the ground may be occluded from the point of view at any selected height. Thus, the problem is to find a set of viewpoints in space from which almost all of the points on the ground are visible. The area formed by the set of visible points so determined is called a *viewshed* in topography [Lee 1991]. The sensor placement problem can now be defined as finding a set of contiguous viewsheds on a terrain or a given wide-area surface, using a minimum set of viewpoints where the sensors may be placed.

Lee [1991] addressed the visibility issues for irregular terrain. He used the Triangulated Irregular Network (TIN) representation of terrain data to find the least number of viewpoints to cover an entire geographic area. The area coverage problem along with the concept of a TIN is also defined and analyzed in Goodchild and Lee [1989]. Eidenbenz [2002] described some approximation algorithms for terrain guarding, which also require estimation of the largest coverage areas from a minimal set of observation points on a rugged terrain.

In both harbor and airport surveillance applications, one of the basic assumptions is that the sensors are placed at a relatively high point with respect to the point of observation. For example, in the case of harbor surveillance, it is expected that the sensors would be located on a bridge tower so as to observe the traffic on the water on both sides and as far as possible on either sides of a harbor channel. In the airport surveillance application, the video cameras are located at strategic points above the usual array of ticket booths, refreshment areas, doorways, etc., to capture images of the traveling public on either side of a camera as far away as possible from the camera, once again to get the largest coverage area. In both applications, the parameter that we are trying to determine is the height of the sensor, and the problem is very similar to locating a viewpoint for a GIS or terrain guarding application as described in the previous paragraph. The set of points and angles obtained at a viewpoint would then provide data about the possible motion of a PTZ camera to enable the capture of the complete wide-area coverage.

By the very definition of the word photogrammetry, it seems obvious that in order to measure physical dimensions of objects from a photograph, there must be techniques readily available that lead to capturing wide-area geometries. Thus, the solutions to visibility problems in this field must readily apply to our problem. A literature search revealed that the visibility problem has been addressed in photogrammetry and remote sensing mainly for building accurate 3D representations of land areas (called orthophotos) and Digital Elevation (or Terrain) Models (DEM or DTM). The following paragraphs briefly survey relevant literature.

Miller [1999] discussed and illustrated the value of digital photogrammetric techniques that help in producing a census of visibility and characterization of landscape views. Maas [1998] discussed close-range photogrammetry that has been used extensively to measure the 3D coordinates in a large number of industrial applications. Using high-resolution solid-state cameras, redundant imaging, geometric and stochastic modeling, and self-calibrating bundle adjustment techniques, accuracies of 1 in 100,000 for object dimensions have been achieved. Murai et al. [1998] reported on the development of a method called the polygon shift method to generate bird's-eye views of topographic landscapes with shade and shadow. The method uses Boolean operations between an original polygon and a shifted polygon with a certain given direction and amount.

Nielsen [2004] experimented with different heights and different wide- and narrow-angle lenses in his study of production of orthophotos. According to Nielsen, the relief displacement (radial distortion introduced by differing relative elevations of objects as

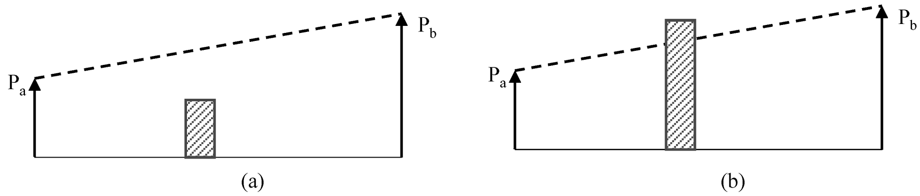


Fig. 2. Intervisibility and blocking object: (a) P_a and P_b are intervisible and (b) P_a and P_b are blocked.

seen from the top by an aerospace sensor) decreases as the flight altitude increases. Poulin et al. [1998] described an interactive system to reconstruct 3D geometry and extract textures from a set of photographs taken with arbitrary or unknown camera parameters. Their basic idea is to let the user draw 2D geometry on the images and set constraints using the drawings. A set of geometric linear constraints formulated as a weighted least-squares problem is then efficiently solved for the camera parameters. Saadatseresht et al. [2004] addressed a sensor placement problem for measuring complex industrial objects by visual metrology systems. They tackled the so-called accessibility uncertainty prediction problem by using fuzzy logic and without having access to any 3D CAD models and workspace information. In Soergel et al. [2001], the authors incorporated a high-resolution LIDAR DEM to investigate the impact of SAR phenomena (like foreshortening, layover, shadow, etc.) on the visibility of the scene. Yastikli and Jacobsen [2003] provided an overview of the automatic generation of DEMs using intelligent filtering and image matching techniques.

Unlike robotic inspection applications wherein the sensor moves around the object in close proximity, the height of the sensor with respect to the observed object in the case of wide-area surveillance is fairly remote and farther away. Furthermore, to get the widest possible coverage area, the sensor's relative height from the object is much larger. Thus, at least two sensor parameters in a wide-area surveillance application, namely the height and the resolution of the sensor, may be selected using the techniques from terrain planning and operational research. The rest of this section summarizes the analysis of visibility [Lee 1991] and the wide-area coverage problem [Goodchild and Lee 1989].

Lee [1991] defined two points on a topographic surface being intervisible if it is possible to connect the two points with a straight-line segment without intersecting any part of the surface other than the two points (see Figure 2). A visibility function is usually defined as a Boolean function, $V_{ab} = 1$, when a point P_a and another point P_b are intervisible.

DEMs are a common representation used in imaging, graphics, and GIS for representing terrains. However, if we would like to denote the vertices of a rectangular grid on a DEM by the Boolean visibility function defined earlier, it would be difficult, as the four general grid points on a DEM may not necessarily represent a plane and it would be an oversimplification to combine the visibility function of the four grid points and use it as a visibility map for the grid. To overcome this difficulty, TIN is the preferred representation for terrains in GIS. A TIN model approximates a topographic surface by connecting irregularly spaced elevation vertices into triangular facets, thus forming a 3D triangular mesh-like model. In this representation each vertex of a triangle facet can be assigned a Boolean visibility function as defined before and used in the visibility analysis.

Given a TIN representation of a terrain (topographic surface) with N points, the i^{th} point, denoted by $P_i = (x_i, y_i, z_i)$, the presence or absence of a sensor at the i^{th} point by C_i, and the area of a subregion formed by placing some kind of a grid over the TIN and splitting it up into parts as A_j, we can formulate the visibility problem in three ways, as follows.

1. For a given set of viewpoints *VP*, find the area of visible regions $\sum_{j \in VR} A_j$ where
$$VR = \left\{ j \,\middle|\, \sum_i V_{ij} C_i \geq 1 \right\}, \tag{1}$$
with $C_i = 1$ if i belongs to *VP* and $C_i = 0$ otherwise.
2. Minimize the number of facilities (sensors) required to see the entire surface, that is,
$$\min \sum_i C_i \quad \text{subject to } C_i = \{0, 1\} \text{ for all } i \text{ and } \sum_i V_{ij} C_i \geq 1 \text{ for all } j.$$
3. Maximize the area covered by a given number of facilities (sensors) F, that is,
$$\max \sum_j A_j \min \left(1, \sum_i V_{ij} C_i \right) \text{ subject to } C_i = \{0, 1\} \text{ for all } i \tag{2}$$
$$\text{and} \sum_i C_i \leq F \text{ for all } i. \tag{3}$$

Lee described some extensions to the aforesaid visibility problems (e.g., he computed a cost function to see the entire surface at some minimum height of a tower at which the sensor may be placed). An interested reader should refer to Lee [1991] for further details and height optimization formulae. It suffices to say here that this is an important paper for ideas and directions to formulate the sensor placement problem for harbor and airport surveillance applications.

3.2. Feature Detectability Constraint

Feature detectability constraints are applicable to both near-field and far-field sensing. Here we are talking about the visible range of an object feature to be detected by an optical sensor. An initial set of constraints for sensor planning was introduced by Cowan [1998] and Cowan and Kovesi [1998]. The constraints depend on whether the illumination plays an important role in the feature detection process. Accordingly two sets of constraints are developed, namely sensor constraints and radiometric constraints.

There are five sensor constraints: visibility, field-of-view, focus, magnification or pixel resolution, and perspective. The *visibility* constraint deals with the problem of occlusion because of obstruction of view by something in the line of sight, be it part of the object itself or a foreign object. The *field-of-view* constraint brings in the requirement that the rays of light to the sensor encompass the entire object so that the object is observable in the active area of the sensor. The *focus* constraint is to guarantee that the image of the object is in focus so that parts of the object may not be blurred. *Magnification* and *pixel resolution* deal with, respectively, the size of feature images and the ability to distinguish two object features to be imaged onto distinct pixels. The *perspective* distortion brings in the loss of depth in the image and is one of the most difficult constraints to deal with, especially, in case of outdoor objects and far-field imaging.

Illumination constraints are illuminability, dynamic range, and contrast. *Illuminability* concerns the ability to illuminate all parts of the object equally well so that the feature on an object is completely visible not only to the sensor, but also to all points on the source of illumination itself, to make sure that the feature is lit in the first place for the sensor to detect it. In addition, if there are any back-facing features to the illumination source, then there will be shadows cast on parts of the object, giving rise to new constraints of shadow avoidance. The *dynamic range* of the sensor depends on the irradiance of the object to the light source. The sensor may interpret a corresponding pixel as being too bright or too dark, thus making the feature undetectable. Finally, *contrast* deals with the ability of the sensor to accurately detect the edge features of an object. This constraint brings in more challenges when the illumination source causes shadows and reflections, both of which contribute towards the detection of false edges.

In a recent survey paper oriented specifically towards 3D object reconstruction, Scott et al. [2003] identified requirements that are a little more stringent and restricted to certain types of sensors, such as laser range cameras and an environment of machine inspection on a factory floor. They call these requirements as view planning requirements and categorize them into general, object, sensor, and positioning system categories. In addition they also place certain set of constraints and further base their planning strategies on a set of assumptions. Further details of this survey are not mentioned here as they do not address the nature of the wide-area sensor planning problems that we are dealing with. However, a basic reading of this survey paper exposes the reader to available techniques and parameters used in sensor planning activities.

3.3. Sensor Positioning in Wide Areas

In order to cover a wide area for surveillance, there are two obvious approaches available: (1) build a network of cooperative fixed pose sensors; (2) have sensors (perhaps omnidirectional) mounted on a mobile autonomous vehicle. In addition, a mobile robot might cooperate with a number of other mobile robots in a network, thus forming a mobile sensor network [Tan and Xi 2004].

Sensors for wide-area surveillance may be positioned anywhere depending on the particular area to be monitored. For autonomous and continuous wide-area surveillance, the sensors need to be of fixed pose. Fixed positions might be somewhere close to the ground on tripods, telephone poles, building corners, etc. Mobile positions could be on top of an automobile driven through the region of surveillance, or on an airplane or a satellite that flies over the region to be surveyed. A second option for fixed pose sensors would be on specially constructed towers such as radio and cellular antenna or fire monitoring towers.

Video surveillance, for instance, has come a long way from the early days of using fixed surveillance cameras placed in known sectors of a building, a parking lot, or the length of an airport terminal and employing human operators to watch a bank of video monitors. The state-of-the-art may not only point to the availability of ubiquitous wireless transmission of the video stream to a World Wide Web server, but also to interactive manipulation of the surveillance camera by the end user watching or monitoring the video. Each of the previous options is discussed in the following subsections.

3.3.1. Ground-Based Fixed Positioning. A majority of video surveillance systems that are commercially in use employ ground-based positioning of their sensors. Usually, video cameras are located at overhead positions of approximately 3 to 4.5 meters above the ground, covering typical living areas like rooms, hallways, etc. [Kang et al. 2001]. Generally, the area of coverage is limited to a few square meters, but it is possible to place a network of ground-based sensors in a wide area with possible communications among the sensors and a central base station [Mhatre et al. 2004]. In certain military applications (naval application, for instance), networks of ground-based sensors are also possible under and on the surface of the water in an ocean. Some examples of ground-based video sensors are found in VSAM [Collins et al. 2001], where a parking lot is being monitored and the sensors are placed on roofs of adjacent buildings, and DIVA [Trivedi et al. 2002], where traffic getting in and out of an interstate highway

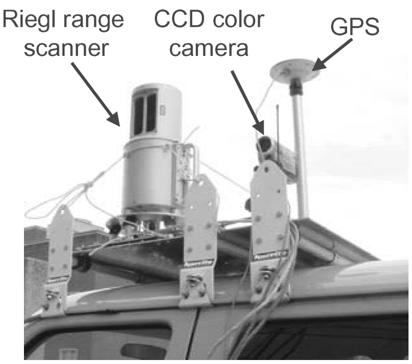


Fig. 3. An examples of mobile vehicle-mounted dynamic sensors, including a Riegl range scanner (left), a CCD color camera (middle), and a differential GPS sensor (right).

in San Diego is monitored with sensors placed on the median and the sides of the road.

3.3.2. Tower-Based Fixed Positioning. Tower-based sensors (also referred to as sentry towers) are mostly used in cellular and radio communication applications where a preplanned wide area of coverage is expected of the sensor. Of late, however, towers are being proposed for active sensing in surveillance applications by the military in securing harbor, large public parking lots, and airport areas. Tower-based sensors [Papasin et al. 2001; Storch 2004] are expected to be engaged in not only active sensing, but also to sensor data communication to central fusion points and public networks like the Internet. In this article, we are mainly interested in situations where fixed pose sensors are mounted on ground/water-based man-made structures that enable long-range and wide-area imaging.

3.3.3. Dynamic Sensors on a Mobile Vehicle. Having sensors mounted on a mobile vehicle for a surveillance application in a wide area has been feasible and utilized by the military in a battle situation [Cook et al. 1996]. Like the static sensor network situation, while dynamic sensors do cover a wide area, they by themselves are not sufficient to form an end-to-end surveillance system. Most of the rest of the system, like central processing, data display and command, and control of the sensors, should come from separate central nodes. However, from a purely surveillance data collection viewpoint, having dynamic sensors, in motion (e.g., in a 3D reconstruction application where sensing is required from all directions) around the object of interest is extremely cost effective and efficient (in that there would be no need for multiple sensor usage and/or calibrations).

An example of dynamic sensors on a mobile vehicle is depicted in Figure 3. They show a mobile platform with a Riegl scanner, a CCD camera, and a GPS sensor mounted on the roof of an SUV used in a research project that involves fast 3D digitization of real-world environments for immersive computer applications [Grinstead et al. 2004]. Yet another example of mobile sensors is found in Miura et al. [2000] where the sensors are positioned on top of a moving vehicle to analyze and recognize traffic signs.

In the context of harbor surveillance for situational awareness, dynamic sensors located on a small coastguard are capable of motion around a large ship entering a harbor. This provides an easy and effective solution to quickly monitor all the activity and goods on the incoming ship in transit, without the long-range imaging complexities that would be introduced by the alternative static sensor towers at the extremities of the harbor.

Table IV. Ground Resolution Required for Targets under Aerial Surveillance (reproduced from, OTA [2003])

Object	Detect (feet)	Recognize (feet)	Identify (inches)	Analyze (inches)
Missile	3	2	6	1.5
Vehicle	5	2	6	1.5
Nuclear weapon	8	5	12	0.5
SSM Site	10	5	6	1.5
Aircraft	15	5	6	1.5
Submarines	25	15	6	1.0
Troop units	20	7	24	6.0

In any dynamic sensor situation, data communication to a central node introduces some challenges, unless the whole surveillance system can be realized within the confines of the moving vehicle. It is obvious that the sensor data communication must be achieved through wireless means and hence all the limitations of wireless communications such as bandwidth, line of sight transmission, excessive channel noise effects, etc., need to be dealt with. Nevertheless, the state-of-the-art technology has matured to a point that miniature and low-power wireless sensors in multiple modes [Crossbow 2008] are becoming available to enable the building of wireless sensor networks with either static or dynamic sensors.

3.3.4. Airborne Mobile Positioning. Sensors in photogrammetry and remote sensing are usually airborne or positioned in geo-synchronous satellites and they usually cover a wide area on the ground. Due to their long-range sensing modalities, these sensors are usually of infrared or thermal type and the sensing operations are usually planned and scheduled to happen at relative positioning of the sensor with respect to the object being sensed or surveyed [Crebolder et al. 2003; Dayton et al. 2000].

Crebolder et al. [2003] reported the use of an airborne, multisensor imaging system for search and rescue applications, called ELVISS (Enhanced Low-light level Visible and Infra-red Surveillance System). They used laser illumination to enhance visibility of an optical camera aboard an airplane and reported identification of more targets with a field of view of 20°. A study commissioned by the U.S. Congress, Office of Technology Assessment (OTA) has resulted in a technical report that details airborne sensor platforms, issues, and operational concerns when using them in surveillance applications [2003]. We summarize its findings in the following paragraph.

Cooperative airborne surveillance may be performed using airplanes, helicopters, and unmanned aerial vehicles (UAVs). The sensors deployed on board might be air samplers or sniffers, radiation detectors to detect chemical and radiological emissions, signal intelligence, passive optical and acoustic sensors, multispectral sensors, laser and radar sensors, magnetic anomaly detectors, etc. Operational considerations include the number and duration of flights relative to the area, composition of the overflown territory, and the amount of advance notice given. Table IV, reproduced from OTA [2003], gives a summary of typical ground resolution required to identify targets as found by the OTA study. As seen from the table, as the distance between target and sensor decreases, the process of figuring out what a sensed target is goes from a mere detection (e.g., vehicle) on one end of the scale to the actual analysis (e.g., a 2004 red cadillac STS with sun roof and license tag TN 1234 GOV). The intermediate steps may recognize the target as being a car (not a truck or a van) and identify it as a 4-door family sedan (not a coupe or a convertible). This kind of increased sophistication with which enough semantic information is gathered at the time of raw sensing would indeed lead to contextual understanding of the scene and more accurate assessment of the situational awareness.

4. SENSOR INTEGRATION FOR WIDE-AERA SURVEILLANCE

With the sizes and complexities involved in most wide outdoor scenes and the importance and sensitivity of many of the tasks required, it is almost impossible for any single sensor to perform any of the surveillance or monitoring tasks with an acceptable degree of continuity or reasonable accuracy. Multiple sensing modalities, with variable attributes and capabilities, need to enter into play. These sensors need to be able to communicate and optimize their energy for the most efficient cooperation to achieve the task required of them.

Many times, sensors used in smart networks are of fixed viewpoint, PTZ, or omnidirectional cameras mounted on computer controlled camera heads. Smart wireless sensors, on the other hand, can be of thermal or acoustic type, collecting temperature and audio signals for transmission to a collection node within the sensor network. Two recent papers point to the adoption of decentralized sensing and tracking [Crespi et al. 2004] and the control of surveillance sensors [Luts et al. 2004]. Crespi et al. [2004] presented a fully autonomous and decentralized surveillance system for detecting and tracking mobile unknown ground vehicles in a bounded area. By contrasts Lutz et al. [2004] discussed the technology of high-performance positioning equipment for pan-and-tilt control in PTZ sensors.

A representative sample of work in the area of smart sensor networks is summarized in the following subsection. Given the importance of sensor handoff, a subsection is also dedicated particularly to reviewing the techniques in this area.

4.1. Cooperative Smart Sensor Networks

Static optical sensors cannot continuously visualize a complete object or feature for analysis and recognition due to many factors, such as obstruction by a foreground element, shadows cast on the object, and occluded backward-facing regions of the object. Therefore, when using static sensors, we need multiple images from as many viewpoints around the object of interest as possible, and the problem gets more complex if the object of interest also exhibits dynamics of motion.

On the other hand, when there is no need for reconstructing a 3D object accurately, feature recognition might be performed, as long as the visibility of the feature is guaranteed from at least one viewpoint for small features and a mosaic of the same for large features. It would also be possible to fuse the optical sensor data with other sensor data such as thermal to recognize a particular feature [Aguilar et al. 1998] (e.g., a running motor with a distinct thermal signature combined with a 2D projected image of the scene to which the motor belongs).

Thus, multimodal sensing along with fusion of sensed data goes a long way in the analysis of the image of interest in a surveillance application. In a similar vein, while the dynamics of an object under observation may impede the process of accurately capturing an optical image (because of the artifacts of motion blur in the image), the very fact that the object is moving might indicate a significant feature in the scene to be analyzed for occupancy or activity. Intelligent systems should make use of every opportunity a sensed image provides to logically infer facts about an object or scene put under surveillance and this makes long-range imaging for surveillance applications different. The challenge posed, however, is in the provision of the required *intelligence* in terms of a trained human being, an autonomous robotic agent, or some combination of both.

One of the strategies used when employing a network of static sensors is to use them cooperatively, as seen in the VSAM system developed at Carnegie Mellon University (CMU) in the late 90's [Collins et al. 2001]. The VSAM system uses *smart* sensors in that each sensor is capable of independent operation and could be added or removed without affecting the others in the network. The cooperation among sensors is achieved by a cost-based sensor tasking algorithm that arbitrates the control among sensors. During the task arbitration of the sensors, handoff between sensors is designed to happen seamlessly according to the algorithm; this is facilitated by rigorous task description and the command language implementation used in the system. For example, in VSAM it is possible to control the smart sensors by such commands as "Track this human" or "Track this van," as opposed to "Move Sensor 1 forward by 0.3m" or "Zoom sensor 2 by 3x," etc. Other cooperative sensing technologies are described in Cook et al. [1996], Matsuyama [2002], Porikli [2004], Ukita and Matsuyama [2002], Wang et al. [2003], and Zhao et al. [2002].

Margi [2003] and Akyildiz et al. [2002] presented thorough surveys on networking, sensor processing, and system aspects of sensor networks. Matsuyama [2002] and Ukita and Matsuyama [2002] discussed the use of active software agents to cooperatively track multiple targets in a sensor network. In Mhatre et al. [2004], we find a division of labor among the sensors of different types (0 and 1) to collect, fuse, and communicate the signals to different cluster points or a central service node. In the DIVA system [Trivedi et al. 2002], sensor communication controls the turning on and off of a specific camera for tracking the current object. Wang et al. [2003] described research on adaptability in surveillance systems by continuously controlling the camera parameters (pan and zoom) and directly communicating the control signals to the sensor. In Zhao et al. [2002] we find a computation of the communication cost (information-driven) for a specific sensor, which in turn is utilized in determining the role it will play in signal data routing in the sensor network.

In summary, the strategy required when using multiple static sensors cooperatively is in the ability of each sensor to handover control or to accept control from another sensor. It is also essential that the sensor task be modeled and described a priori in the system in an elaborate fashion. For a harbor surveillance application, for instance, with four sensor towers, each carrying multiple multimodal sensors, sensor tasking and control would be impossible without an extraordinary effort in task modeling.

4.2. Sensor Handoff

The concept of a handoff during tracking follows naturally when continuous tracking of a moving object is desired in the context of a limited field of view of any one sensor. Handoffs are also significant when an intelligent surveillance system determines a need to focus on some activity within a narrower view of a wide-area scene by zooming onto a specific region within [Kang et al. 2002, 2001]. Yet another context for a handoff might be a need to switch the modality of the sensor from color video to infrared or thermal as the subject of tracking has moved into shadows or areas of low visibility. The concept of handoff has been studied more broadly in literature under the so-called cooperative sensing using smart sensing distributed networks [Hampapur et al. 2003]. Sensor networks comprise a multidisciplinary topic [Margi 2003] that involves both networking and signal processing. In Chu et al. [2002], for example, the researchers described two novel techniques (IDSQ: Information-Driven Sensor Querying and CADR: Constrained Anisotropic Diffusion Routing) for energy-efficient data querying and routing in ad hoc sensor networks. Back in the late 90's, the Video Surveillance and Monitoring (VSAM) team at the Robotics Institute of CMU [Collins et al. 2001] developed an end-to-end system that allowed a single human operator to monitor activities in a cluttered environment using a distributed network of active video sensors. The U.S. Navy [Control and Ocean Surveillance Center 1997] sponsored research for developing an integrated acoustic sensor network with an optical docking technology for battery charging and surveillance data transfer.

In Porikli [2004], we find a slightly different strategy to track objects based upon the video data for each object, as opposed to video data from each sensor. They used background subtraction and mean-shift analysis to track each object as it moves in the area of nonoverlapping video field-of-views. Trivedi et al. [2002] described their Distributed Interactive Video Array (DIVA) system as using an event-driven servoing technique to capture desired events at appropriate resolutions and perspectives. They used handoff schemes for passing tracked objects between sensors and clusters of sensors, determining the best view from the scene context and employing the strengths of a specific sensor or its modality while also fusing the data to assist their surveillance task. Adaptability in real time is one of the requirements identified earlier as an essential ingredient of wide-area surveillance. Kang et al. [2001] computed and used the height ratio and motion information to determine the best viewing camera currently available for tracking a person and to accordingly make a decision to switchover control to another camera. Wang et al. [2003] presented a scheme for adaptive monitoring of surveillance objects by utilizing the feedback of the experimental sampling results to change the video camera parameters. Finally, Zhao and collaborators at Xerox PARC developed an information-driven approach [Zhao et al. 2002] to determine participants in a collaborative sensing process by dynamically optimizing the information utility for a given cost of computation and communication. In summary, all these new smart sensing approaches are becoming an important (albeit, expensive) part of wide-area surveillance and monitoring systems.

4.3. Data Fusion

Data fusion has played an increasing role in the use of sensor networks in current research, as presented by Aguilar et al. [1998] whose system creates a night color vision capability based on biological models of the spatial and opponent-color processes in the human eye. The role of autonomy for surveillance of multiple targets in a rectangular zone with a distributed sensor network is studied in Krishna et al. [2004], where sensors coordinate amongst themselves to allocate to a particular target. This type of autonomous behavior would be crucial in wide-area surveillance, where incidents are mostly localized to small regions and it would be important for the sensor to refocus its attention (e.g. breach detection) in real time without operator intervention. Two additional examples of security and traffic monitoring of multitarget tracking in real time using active vision agents are described in Matsuyama [2002] and [Ukita and Matsuyama 2002]. Sensor cooperation and data fusion is further exhibited in Mhatre et al. [2004] where two types of sensor nodes are used for data collection and dissemination. Type-0 nodes perform sensing and relaying of data within a cluster of nodes in a large grid of sensors surveying a wide-area. Type 1 nodes act as cluster heads or data fusion points to fuse the incoming data from type 0 nodes to be relayed to a surveillance aircraft that visits the sensor grid at periodic intervals for data collection.

In intelligent systems, data gathering from sensors is more than just an exercise in deploying the required sensor hardware in known locations and controlling them (either manually or autonomously) to acquire raw data. Hager [1990] pointed out four main principles that determine a theory of intelligent information-gathering system: (1) task-directed, that is, tailoring the sensing activity to maximize system performance; (2) uncertainty, that is, the decision as to what feature under surveillance is being sensed is always uncertain; (3) computational limitations, that is, gathering irrational amounts of information (sensor data) without bounds is not optimal and would require infinite amounts of computing resources; and (4) representational limitations, that is, unless the sensed data may be converted into a useful form amenable to some efficient semantic interpretation, the sensed data would be pointless.

Hager [1990] succinctly explained information gathering or sensing in terms of a state diagram where the final goal state is to obtain an adequate description of particular aspects (in long-range imaging for surveillance, this would be a well-known threat feature!). Hager went on to define the fundamental aspect of this purposeful information gathering as being dependent on the type, accuracy, and cost of information processing needed, and called this task modeling. A task modeled information gathering and the process that combines and evaluates the sensor observations is central to the problem of information gathering in intelligent systems and is defined as task-directed sensor fusion.

Further characterization of the details of task modeling and sensor fusion at this juncture is out of the scope of this article. However, Hager [1990] takes up the discussion and description of the sensor fusion theory, mathematics, and real-world examples in more detail and the interested reader should refer to Hager's seminal work on task-directed sensor fusion.

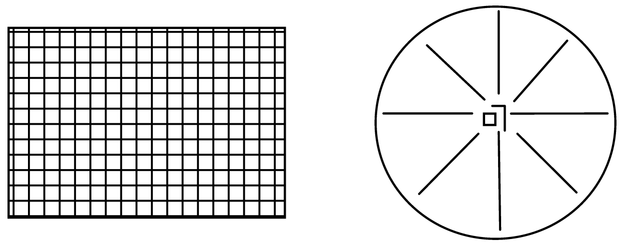
A landmark survey paper regarding multisensor data fusion is given by Hall and Llinas [1997]. Another more recent survey paper by Smith Singh [2006] added newly proposed algorithms in this area with a focus on target tracking applications. In Smith and Singh [2006], data fusion is divided into four subtasks: data registration, data association, position/attribute estimation, and identification. Various approaches are discussed for each subtask. For instance, nearest-neighbor, joint probabilistic data association, Lagrangian relaxation, artificial neural network, and fuzzy logic approaches are introduced for data association, while Kalman filtering, multiple-resolution filtering, particle filtering, artificial intelligence, and fuzzy logic adaptive filtering are reviewed for position/attribute estimation.

The primary source of *intelligence* in surveillance systems is comprised of the visual data as well as the analytical abilities and past experiences of a human to interpret the visual data being captured. However, the perspective distortion experienced by the visual sensors brings more difficulties to the fusion of visual sensors. Therefore, in the following sections, we concentrate on the fusion of visual sensors.

Visual input considers the following essential parameters [Nicolescu et al. 2000]: region of awareness, region of interest, background model, depth range, and cost. In long-range imaging, the region of interest is wide by necessity and the region of awareness depends on the placement and number of sensors. Mosaicing techniques are concerned with creating the required large region of interest from the available limited regions of awareness. One of the primary means of gathering interpretable visual data using long-range observations is by efficiently stitching together pieces of connected high-resolution data that may be present in a specific location on the grid over which imaging is being conducted, as a planar, spherical, or cylindrical mosaic.

Mosaicing techniques for the creation of very high-resolution images and panoramas have garnered a lot of interest in recent years with the advent of the World Wide Web as a preferred medium of publishing and such advances as the virtual reality formats by academic and industrial researchers. The literature points to a wide variety of amateur and professional systems and techniques for creating large panoramas [Peleg and Ben-Ezra 1999; Peleg and Herman 1997; Peleg et al. 2000; Rousso et al. 1998; Brown and Lowe 2003; Capel and Zisserman 1998]. A comprehensive treatment of all the techniques and systems is beyond the scope of this article. However, a review of some basic techniques, formats, and challenges is given in the following paragraphs.

There are two formats in which a high-resolution image may be constructed by stitching together a number of lower-resolution images. They are: (1) planar mosaicing, where individual images from translational motion cameras are stitched together horizontally and/or vertically, as shown in Figure 4(a); and (2) a panorama rendered as a planar, spherical, cylindrical, or cubic projection, where the images surrounding a fixed camera



(a) (b)

Fig. 4. High-resolution mosaic and panorama: (a) planar mosaic from a grid of low-resolution images; and (b) spherical panorama from images taken in angular sectors.

in angular increments (horizontally and/or vertically) are stitched together, as shown in Figure 4(b).

In long-range surveillance applications, a planar mosaic employing a grid of element size (10m×10m when covering 10km×15kilometer area, for example) would be desirable. Szeliski [1996] discussed the mathematical algorithms involved in the creation of planar mosaics and spherical (or cylindrical) panoramas and gave some examples of creating simple mosaics of short- and long-range scenes. Essentially, there are two methods by which it is possible to stitch together individual images taken over a grid or angular sectors, as follows: (1) manually compositing adjacent images by overlaying one over another so that two or more known features of the images coincide, with the assumption that the adjacent images overlap to some known extent (say 10% or 30% along the horizontal or vertical axis); (2) automatically registering two images by using some registration algorithm that uses intensity of pixels averaged over a known area or other available alternatives.

Manual methods, while being slow and tedious, are less expensive to implement than automatic methods, which may need expensive computing resources and calibrated camera equipment. Further detailed discussion of algorithms for stitching will be deferred and the interested reader is referred to the relevant papers listed under References. For an excellent survey and comparison of image stitching techniques, the reader is referred to Chen [1998].

Individual images from multiple sensors may be fused into large mosaics that could be rendered in planar, cylindrical, or spherical projection, thus reconstructing the complete object under observation (perhaps even in 3D, given sufficient compute power and time) in high resolution. However, if the speed of the moving object in the scene is beyond a given threshold, the reconstructions will be inaccurate and blurry. In case of wide-area surveillance of a harbor with slow-moving ships and barges, it is expected that the reconstruction of objects of interest (i.e., large ships and barges) from video mosaics would be satisfactory for any potential threat characterization and situational awareness. Figure 5 illustrates an example of a planar mosaic via stitching of selected video frames using the Realviz Stitcher.[2] The barge video data was collected in high-definition format by a 6 Megapixels JVC digital video camera mounted on a tripod on the banks of the Ohio River in Cincinnati, Ohio.

Super-resolution (SR) can be considered as the opposite type of operation when compared with image mosaicing. Image mosaicing takes in images with plenty of detailed contents of a small FOV and generates an output image with an enlarged FOV while maintaining the same detail level as the input images. In comparison, the input images of an SR algorithm already have a large FOV but lack the desired details or resolution.

[2]http://www.realviz.com/.

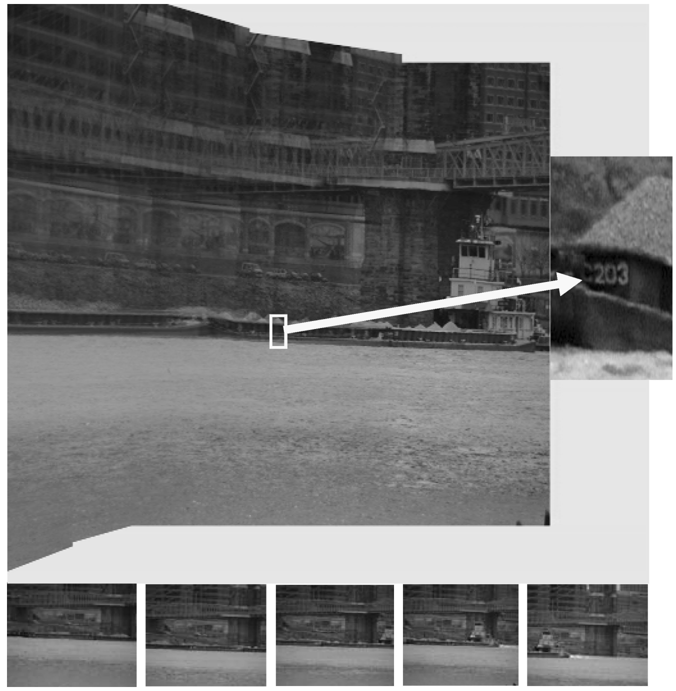


Fig. 5. Example of a planar mosaic (top) from five high-definition video frames (bottom); the inset shows a magnified view of some tag numbers on the barge.

The goal of SR is to construct an output image with the same FOV and improved degree of details or resolution. As the counterpart of image mosaicing, SR techniques also address the balancing of region of interest and region of awareness. SR has been receiving intensive research interest since the pioneer work of Tsai and Huang [1984] and abundant SR algorithms have been proposed [Lertrattanapanich and Bose 2002; Bishop et al. 2003; Vandewalle et al. 2005; Wagner et al. 2004; Marcel et al. 1997; Keren et al. 1988; Ferreira 1994; Irani and Peleg 1991; Alam et al. 2000; Ur and Gross 1992]. With limited optical zoom capability restricted by the system's hardware configuration, SR algorithms provide a promising solution with no additional hardware requirements. Borman and Stevenson [1998] reviewed existing SR algorithms. A more detailed and updated review can be found in Park et al. [2003].

An SR algorithm usually consists of two major steps: registration and interpolation/reconstruction. In registration, the displacement among input frames is estimated. Based on the estimated transformation, all low-resolution images are aligned and a high-resolution image is then generated by interpolation or deconvolution. Marcel et al. [1997] and Vandewalle et al. [2005] explored the relation between spatial translation/rotation and magnitudes/phases in the frequency domain to estimate a 2D interframe transform and register consecutive low-resolution images. A direct spatial-domain implementation makes use of the Taylor expansions [Irani and Peleg 1991]. The Papolis-Gerchberg algorithm is the pioneer interpolation algorithm for SR [Ferreira 1994]. More recently, the constraint least-squares algorithm has been proposed, which is actually a deterministic regularization method for solving the inverse problem from the observation model [Park et al. 2003]. This approach has become the dominant and most popular approach for interpolating low-resolution images.

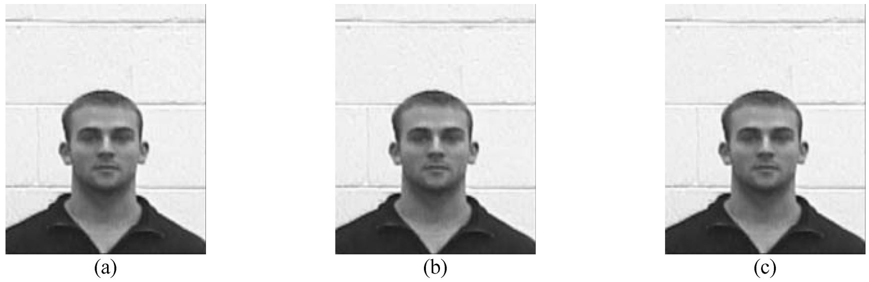


Fig. 6. Illustration of SR images: (a) original image shown at 60% of the original image size; (b) and (c) SR images with a SR factor of 2 and 4 shown at 30% and 15% of the original image size, respectively. The SR factor is defined as the ratio between the height/width of the output high-resolution and the input low-resolution images.

SR algorithms have been successfully used in long-range face recognition. Most existing face recognition engines require that the input face image have a minimum pixel count between the eyes to ensure a successful recognition. It is shown that face recognition rates drop gradually with respect to the decreased pixel count or, equivalently, image resolution [Yao et al. 2007]. SR has become a method of choice for increasing the resolution of a face image and in turn the resulting face recognition rate. In addition to general SR methods, algorithms designed particularly for face images also exist in literature, such as hallucinating faces [Baker and Kanade 2000] and Eigenfaces [Gunturk et al. 2003]. Baker and Kanade [2000] employed a pyramid based algorithm to learn a priori of the derivatives and then reconstructed the high-resolution images. A layered predictor network is employed for face SR based on a resampling-maximum likelihood model [Lin et al. 2005]. Jia and Gong [2005] addressed the reconstruction of SR face images using multiple occluded images of different resolutions, which is commonly encountered in surveillance videos.

Figure 6 illustrates an example of super-resolved face images where the interocular distance increases from 35 pixels to 60 pixels (the minimum requirement of FaceIt, a well-known face recognition engine [Phillips et al. 2002]), and further to 140 pixels.

4.4. Object Characterization and Recognition

The final goal of data integration in wide-area surveillance is the identification of a threat object or threat situation. Buxton [2003] provided a comprehensive review of the state-of-the-art research in learning and understanding dynamic scene activity. Image understanding and learning are essential prerequisites to object recognition. The task of recognizing an object in order to predict a future action such as a potential terror threat is indeed difficult. This whole task would be dependent on some known premise or an intimate knowledge of the task at hand. Thus, some researchers have embarked upon task modeling [Hager 1990] before attempting any object recognition from a set of image data.

One of the emerging themes in object recognition research is the use of multisensor data and data fusion. Typically, color video or infrared (thermal) data is used as a texture overlaid on the reconstructed 3D data from range scanners. This makes the overall scene segmentation in an image much easier to render when aided by a human operator in an interactive mode. Several recent papers have explored recognition in outdoor environments using active sensors, concepts from AI in analyzing/learning the activity in a scene, multiple levels of zoom to focus on parts of a scene, etc., to contribute towards building intelligent surveillance systems. In addition, paradigms from remote sensing and photogrammetry are being merged into computer vision to enable a broader approach to intelligent monitoring of large public areas.

One of the techniques used in recognizing objects in a scene is by using neural nets with training sets and progressive learning. In Crane [2002], we see a comparison of two approaches, kernel expansion and transductive support vector machines, as applied to an active vision task involving vehicular traffic. The authors' motivation is to learn to recognize objects from very few labeled training examples (i.e., to infer information about an unknown object quickly and without much training). De la Escalera et al. [2003] dealt with a similar situation where they tried to recognize traffic signs in poor lighting conditions and in the presence of other occluding objects. They used a genetic algorithm for detection and a neural network for classification. Yet another real-time traffic-sign recognition system is documented in Miura et al. [2000]. Forsyth and coworkers [2002] argued that any computational theory of object recognition should explain how to decide and recognize objects at the level of abstract categories. Their thesis is that perceptual organization, *not* object primitives representation, is the key in inferring objects from images. Learning and adaptability in vision research is further evidenced in the work of Heisele [2003] where a component-based approach is utilized. Also, Zlatanova and Tempfli [2000] presented an integrated approach to 3D modeling of complex scenes of deterministic objects. This paper would be of importance to wide-area surveillance in terms of spatial scene analysis for object recognition by modeling visible geometric primitives such as points, lines, surfaces, and body, and their topological inter-relationships. Finally, in Smith et al. [2004], we find an interesting approach employing multiple levels of zoom in active vision research for recognizing human activity. Here the researchers make several uses of the epipolar constraint in the context of activity recognition to show that by utilizing multiple levels of detail, we are able to answer recognition problems that a single level of detail may not allow.

Thus, the task of object recognition in wide-area surveillance can be broken down into the following distinct steps [OSUP Lab 2004]: (1) raw multisensor data collection and processing, (2) segmentation, (3) feature extraction, (4) semantic grouping, (5) model fitting to a known object, and (6) model recognition. In using multisensor data for object identification, the raw data might have to be fused at three levels: data level, feature level, and decision level. While the benefits of using multisensor data are increased confidence and reduced ambiguity, problems remain in terms of increased complexity and challenges in interpretation without an accurate a priori knowledge of the model on which the recognition would be based.

5. SYNTHESIS AND RECOMMENDATIONS

In this section we synthesize strategies and put forth recommendations that we believe are pertinent to a realization of pragmatic and practical wide-area surveillance systems. While sensor and computer vision technologies have come a long way since the late 1980's and the need for smart surveillance systems has increased significantly in the wake of the September 11, 2001 tragedy, it is still complex and costly to build and deploy purely autonomous systems to meet security needs. However, a smart network of cooperative, semi-autonomous sensor systems may be designed and built for specific surveillance tasks using guidelines and strategies as outlined in the following paragraphs.

5.1. Long-Range (Wide-Area) Sensing Requirements

The requirements for long-range imaging in surveillance applications are markedly different from those for other applications in which the area of coverage is limited, as

Table V. Wide-Area Surveillance Requirements

Item No.	Item of Interest	Requirement/Characterization
1	Image resolution	Variable – low, medium, and high as required.
2	Accuracy/tolerance	Centimeter to meter accuracy is sufficient.
3	Object dynamics	Objects under observation are mostly dynamic; moving at different velocities and directions; but the sensor locale is fixed.
4	Object size/coverage area	Differing sizes of objects in outdoor environment; large surface area coverage (both in horizontal and vertical planes are required.)
5	Near or far-field object	Generally, far-field object with viewing distances from tens of meters to a few kilometers; requires zooming in to centimeter resolutions.
6	Number of sensors	At least two, for instance, just for day time only, one wide angle and one PTZ sensor; could use a network or array of cooperating sensors for tracking purposes.
7	Modality of sensors	Stereo/color visual, laser/radar range, thermal/infrared.
8	Wide area tasks	Motion detection, tracking, motion classification, feature selection and extraction, object classification, etc.
9	Application areas	Military/civilian surveillance, Photogrammetry, Terrain Planning, and Remote Sensing.

is the case of machine inspection of manufactured parts, reverse engineering for 3D reconstructions, and other near-field vision applications where the sensor pose is fixed relative to the scene. There are many unique sensing requirements in the case of wide-areas. The resolution at which an object in a scene is to be sensed is dependent on the object and could be low, medium, or high. It would be low when the overall scene is sensed, but higher when a particular object within the scene is zoomed in for a better observation. Similarly, the accuracy with which a given sensor is sensing data does not need to be absolutely high. Unlike the application of an inspection vision system where the manufacturing tolerance is required at millimeter precision, the sensed data in wide-areas might only need to be accurate to a degree of a few inches or centimeters. More importantly, the object under surveillance is not static in wide-area applications. There will be object motion, relative background motion, and also dynamic sensors (e.g., PTZ camera). The motion dynamics and the associated complexities for object detection and tracking make it harder in a wide-area application to plan sensor placement and collect sensor data.

The object size and coverage areas differ in a wide-area application, thus adding the complexity of data fusion via mosaicing techniques and fusion of data from multiple sensors. The viewing distances may be tens of kilometers and zooming in on the object may give rise to multiple images of the same object in differing sizes, orientations, and resolutions. High-magnification issues and defocus are some of the other aspects to be dealt with while sensing from a long range. In addition, multiple modalities of sensors may be employed and data fusion from visual, range (laser or radar), and infrared (thermal) is to be expected in a wide-area surveillance application. Applications identified in wide-area sensing are military and civilian surveillance, photogrammetry and remote sensing, terrain guarding, and forest fire fighting. Table V shows items of interest and their corresponding requirements, reflecting the challenges in long-range imaging.

5.2. Zoom vs. Focus

The resolving power of an optical instrument or system is the ability to separate two closely spaced objects. Two extreme examples are separating two cells of a diatom under a microscope or separating a double star with a telescope! The resolving power is also known as the *resolution* and for 2D images a horizontal and a vertical resolution are separately identified.

The fine detail resolved is expressed in terms of the highest spatial frequency in cycles per millimeter (unit mm^{-1}) which can be separated visually in an image. In digital imaging, the resolution is expressed in terms of pixels per millimeter, either horizontally or vertically. It can also be expressed as the number of pixels an image plane contains horizontally and vertically, but often the two resolutions are multiplied together to form the total number of pixels available in an image plane. Thus, we hear of a 6 Megapixel camera or an 11.1 Megapixel camera, etc. Since the image plane also has a known aspect ratio (usually 4:3 or 16:9, etc.), we can easily calculate the individual horizontal and vertical resolution of an image supported by a given camera [Ray 1994].

The resolving power S of a given sensor can be expressed as

$$S = \frac{1}{R} = \frac{1}{1.22\lambda N} mm^{-1}, \tag{4}$$

where R is the Rayleigh limit (used as a spatial frequency), λ is the wavelength of the light radiation in use, and N is the f-number for a circular aperture or the effective aperture, as the case may be, for the lens in use. Now, if u represents the distance between the lens and the object and f is the focal length of the lens in use, we can write the minimum object dimension (in millimeters) resolved, L, as

$$L = \frac{uS}{f} mm. \tag{5}$$

Zooming is complex for a variety of reasons and in any long-range imaging (as is the case for wide-area surveillance systems), the use of zoom functions to get the best possible detail of a feature becomes difficult without knowledge of the system of lenses used in zooming, effect of zooming on focus, etc.

Zoomed-in images tend to reduce sharpness or cause blur to increase the apparent resolution at the cost of clarity. To put this in different words, zooming causes defocus or blur. Thus, it is very important that a system using a zoom lens has mechanisms to correct the focus when zooming is performed on an object or feature of interest. In most commercial digital cameras, zooming is achieved in concert with autofocus so that the clarity or sharpness of the image of an object of interest is kept constant across varying zoom factors.

Ray [1994] discussed focusing of optical systems in an elaborate way. What is important for a zoom lens system is some form of autofocus mechanism built into the system. A wide range of methods are in use for autofocus, from simple mechanical systems to complex signal-processing-based systems. However, the basic principle used is to satisfy the lens equation by adjusting the image plane distance to get a fixed focal length f as the object distance is changed while zooming. The following paragraph summarizes the ranging method used in most autofocus cameras.

The basic strategy used is EDM: Electronic Distance Measurement. The range is measured by using any one of the available ultrasonic, infrared, laser ranging methods and adjusting the image plane distance by moving the lens or the light-capturing mirrors behind the primary lens. Other ranging systems scan the subject region to determine the point of focus, indicated by a small outlined area in the viewfinder. These and other autofocus methods used in the industry are listed next and their details are elaborated in Price [2007]: (1) ranging by image contrast measurement; (2) ranging by exit pupil measurement; (3) autofocus by phase detection; (4) wide-area autofocus; and (5) autofocus flash assist.

Digital autofocus is also an area of active research, where Point Spread Functions (PSFs) are automatically determined from data and regularization techniques are employed to digitally focus a blurred image [Kim et al. 2004].

5.3. Sensor Selection and Positioning

For around-the-clock wide-area surveillance, at least two sensor modalities are required, namely visual (optical) and thermal (infrared). For all-weather operation, we need to complement the visual and thermal sensors with radar and laser range sensors. A possible sensor setup would be two visual (one wide angle, i.e., omnidirectional or fisheye, and one PTZ) sensors, one thermal sensor, and one range sensor for a simple 2.5D autonomous detection task. The number and type of sensors would increase dramatically as the requirements get more stringent and more specific in terms of desired results from the surveillance task.

The strategic positioning of sensors for all kinds of surveillance tasks would be impossible without a priori knowledge of specific task requirements. However, for a general wide-area surveillance application, a cooperative network of fixed location and statically positioned sensors that gives the widest possible visibility coverage at all times is recommended. In addition, the data collection, data fusion, and intersensor communication controls (see the next two subsections for more details) need to be exercised via a centralized control node and a human operator interface (available ubiquitously and wirelessly).

5.4. Sensor Data Networking and Data Fusion

Secure wireless data networks with a capability to bridge the Internet are becoming increasingly robust, readily available, deployable, and, more importantly, cost effective for wide-area surveillance applications. The modern architecture of ubiquitous wireless data networks lends itself to incremental modular growth, ease of maintainability, and rapid means by which the network capability may be advanced in terms of bandwidth and availability. Decoupling the data network and sensor control from data processing elements not only increases operational flexibility, but also the amount of intelligence and complexity that could be handled by the large back-end processors located at a central node. In addition, the central node itself might be linked to other sources of surveillance intelligence databases and networks, thus improving the quality of data analysis and generated results. By generating the results in Internet Web formats (e.g., VRML), simultaneous distribution of the results of any security analysis would be possible for a variety of Web client devices such as laptops, handhelds, and even cell phones.

As mentioned earlier in Section 4, sensor data fusion from multiple sensors is needed for generating visual mosaics, adding texture and color to range images, etc. Multiple data streams bring in more information about the situation under observation and increase the quality of data analysis by eliminating such artifacts as false positives. Data fusion also aids in computing better metrics (e.g., length and width) of objects in a surveillance image (e.g., 3D reconstruction from multiple images of an object from multiple viewpoints). Working with fused data increases the computational complexity and the computing power required at a central data processing node. But the increase in computational cost is balanced by less human intervention and more autonomous control, and hence less risk to human lives.

5.5. Semantic Gap and Image Understanding

An exciting and new field to emerge as a logical extension to the World Wide Web is *Content-Based Information Retrieval* (CBIR) and the concept of a *Semantic Web* [Eidenberger 2004; Palmer 2002; Smeulders et al. 2000; Shadbolt et al. 2006].

Researchers have been busy since the late 1990's in figuring out ways to query, search, and retrieve "meaningful images" from the Web, as opposed to a lengthy textual document that describes a given context. The obvious motivation for this is the old cliché, "A picture is worth a thousand words." Likewise, in wide-area surveillance for situational awareness, the current methods of extracting low-level features from images with a good degree of efficiency should be complemented by the extraction of high-level semantic concepts from an image of a developing situation. Smeulders et al. [2000] defined a lack of coincidence between the information that can be extracted from visual data and the interpretation that can be assigned by a monitoring user, as the *semantic gap*.

Narrowing this semantic gap while contributing to image understanding will be an important element of wide-area surveillance systems in the near future. Visual Information Retrieval (VIR) systems using wireless sensor networks and fused image data are becoming increasingly important, as can be seen from a handful of papers in the literature [Chang et al. 1997; Eidenberger 2004; Gupta 1997]. However, as Eidenberger [2004] pointed out, current VIRs are unable to do visual reasoning (a la humans) and recognize the real-world objects behind 2D images. Nevertheless, we feel that as a first step in a semi-autonomous system, visual information retrieval is expected to play an important part in the design of wide-area smart surveillance systems.

We also feel that three trends are essential to building a successful wide-area surveillance system to understand and interpret an impending threat situation [Smeulders et al. 2000]: (1) similarity and learning, (2) user/agent interaction with the scene, and (3) the need for large databases. As Smeulders et al. [2000] have already pointed out, little work cutting across computer vision and database disciplines has been done and this needs to be hastily corrected. Also, while image learning has gained a lot of attention from pattern recognition researchers and while models are now being built using neural nets for classification and similarity or consensus building, we see further need for similarity-induced semantics and relevance feedback in intelligent systems. The next subsection will discuss interaction and relevance feedback in more detail.

5.6. Relevance Feedback, Open-Closed-Loop Control

One of the identified problems in wide-area surveillance is the amount of raw data collected by sensors and the amount of sifting through (or image segmentation) that needs to be done in order to make sense out of a live situation. Researchers in the area of information retrieval using WWW-based search engines face similar problems and have come up with a strategy called CBIR, mentioned earlier. One of the proposals for bridging the semantic gap between detailed (low-level) image features and a user's need for information content in the image is called relevance feedback [Rui et al. 1998]. Relevance feedback can be used in wide-area surveillance in terms of a smart collection of relevant data from a scene. This in turn manifests itself as a sort of closed-loop control of sensing devices gathering surveillance data. This contrasts to the open-loop methods currently used, in which the data collection is largely performed in a 24/7 continuous surveillance mode with forensic data analysis being performed at a later time and after a tragic event has taken place.

The object model O, proposed by Rui et al. [1998] to use the relevance feedback mechanism,

$$O = O\,(D, F, R), \tag{6}$$

where D is the raw image data a JPEG image (e.g., $F = f_i$ is a set of low-level visual features associated with the object in the image) (e.g., shape, color, and texture) and $R = \{r_{ij}\}$ is a set of representations for a given feature f_i (e.g., color histogram and color moments for the color feature).

In general, there might be K components for each r_{ij} and the model supports differing weights, W_i, W_{ij}, and W_{ijk}, associated with f_i, r_{ij}, and components r_{ijk}, respectively. An example [Rui et al. 1998] of differing weights would be in the wavelet decomposition of textures, since the mean of a sub-band may be corrupted by the ambient lighting condition, whereas its standard deviation is independent of the ambient lighting conditions, whereas its standard deviation is independent of the ambierilighting conditions. The goal of relevance feedback is to find appropriate weights that model the user's (or, in the case of wide-area surveillance, the task's) information need.

Relevance feedback may also be used to *learn* in intelligent systems, as evidenced by the MARS2 system developed by researchers at Microsoft Research and University of Illinois at Urbana-Champaign [Rui and Huang 2000]. Learning (often not given enough priority, albeit it contributes to the total cost of developing an intelligent system) is an important aspect of wide-area surveillance systems and should be actively pursued at every facet of system design, especially in the selection and placement of sensors for surveillance data collection.

5.7. Future Directions

We feel that future research in intelligent wide-area surveillance systems is wide open, given the available, relatively low cost of computing hardware, sensing technology, and data communication infrastructure. In addition, the time is ripe for research into stronger object and scene modeling techniques based on specific surveillance tasks, and integration of technologies available from multidisciplinary fields as diverse as remote sensing and information retrieval. Some of the immediate research and experimentation needing attention are: (1) integration of digital video into inexpensive and miniaturized wireless sensors that are deployable in sensor networks; (2) modeling of human abilities of image perception into threat object descriptions; (3) intelligent surveillance data capture based on relevance feedback; (4) near-real-time algorithms for *abstract* object/scene reconstruction to enable better situational awareness; and (5) multimodal sensor data fusion to enable complex models that are able to close the apparent semantic gap that exists in raw sensor data.

Figure 7 illustrates our concept of *Smart Wide-Area Networked Surveillance* (SWANS). It is essentially a wide-area network of cooperating sensor subnetworks, intelligent local data fusion and control points, public Internet and/or other secure private broadband networks, a set of ubiquitous and mobile surveillance clients, and a central command and control center which houses the human interface, knowledge databases, and required programming models for guiding the entire system. We see five conceptual components of a futuristic wide-area surveillance system as follows: (1) cooperative sensor network; (2) regional field nodes for integration and intelligent analysis of multimodal raw surveillance data; (3) available public Internet and/or secure wide-area broadband wireless backbone for communication and ubiquitous control; (4) surveillance access points via ubiquitous wireless devices in varying forms; and (5) centralized command and control.

The cooperative sensor network performs the surveillance tasks of raw sensor data collection and intelligent, relevant sensing while performing the required tracking of suspicious activity via sensor handoffs and autonomous zoomed-in data collection. The regional field nodes are points of intelligent data fusion, data analysis for situational awareness, and a second-level autonomous sensor network control. It is expected that timely, on-demand information is transmitted to and received from the central command for ongoing, 24/7 surveillance tasks envisioned by the central command operators and directives. Broadband, wireless networks (including the Internet) perform the

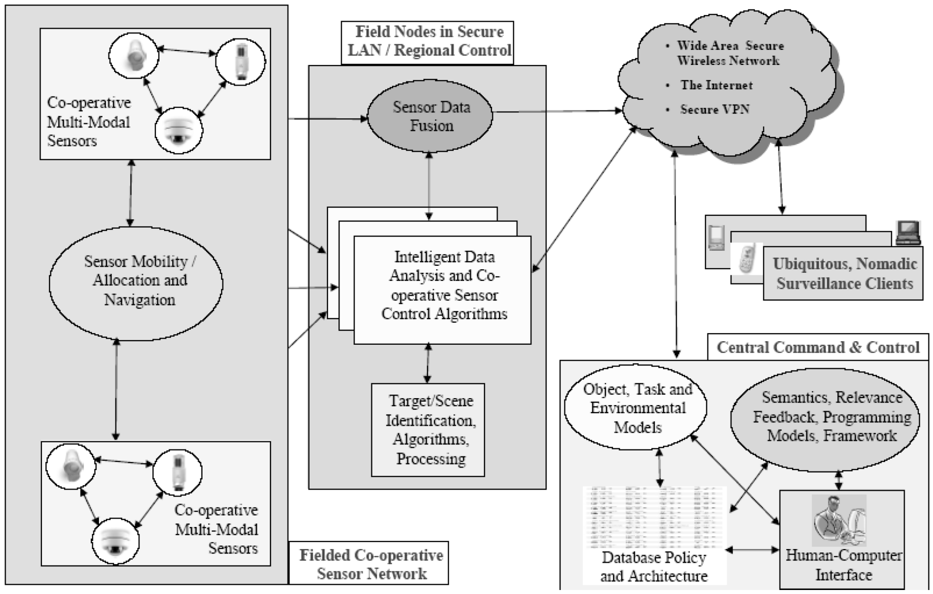


Fig. 7. Conceptual smart wide-area networked surveillance system.

seamless and secure digital data communication required between all the participating elements of SWANS. Individual wireless and/or wired clients access surveillance data on a secure, subscription basis in various levels of access control (i.e., read-only or view-only, read-write or view-monitor, etc.) and with secure login facilities to specific nodes in the network, for the purposes of conveying situational awareness and control information. The central command and control functions as the glue that holds all the individual elements and performs the necessary operational and maintenance tasks of the overall system. Needless to mention, this overall concept is not intended to convey ideas about any existing or planned system per se, but only to reflect the complex nature of this research and its underlying functional ramifications.

On a final note, we would like to identify some ongoing research activities in some of the future directions listed earlier by INRIA of France[3] and the European Commission (EC). These two bodies are collaborating on a project called CAVIAR (Context-Aware Vision using Image-based Active Recognition), whose goals are to use contextual knowledge in gathering surveillance data features through a selective attention mechanism and context models from a database. In fact, they have already published a set of CAVIAR test case scenarios[4] depicting the ground truth via video clips of such human activities as "people walking," "people meeting," "window shopping," "fighting," "person leaving a package in a public place," etc. We feel that the future for research in intelligent surveillance systems in wide public areas is indeed open, exciting, and necessarily multidisciplinary.

6. CONCLUSIONS

A comprehensive survey and analysis of sensor technologies and surveillance systems for wide-area surveillance applications has been studied in-depth. We have presented

[3]www.inria.fr/index.en.html.
[4]http://homepages.inf.ed.ac.uk/rbf/.

an overview of sensor selection and planning requirements, visibility analysis and wide-area sensor positioning, considerations for zooming and rotating sensors, data fusion for salient feature detection, and a synthesis of recommendations.

Wide-area surveillance has become a research topic of choice among computer vision, electrical engineering, and computer science researchers in both academia and industry, answering to an increased need for around-the-clock, autonomous surveillance requirements by the government and the security industry. The problem at hand is still complex, expensive, and experimental at the time of this writing. However, newer advances in the state-of-the-art computing hardware, networking, and software algorithms keep pushing the research techniques higher and farther. We have discussed advances in intelligent sensor networks, data fusion, sensor placement, and dynamic sensors which make possible to construct 24/7, ubiquitous, active surveillance and monitoring systems to ensure homeland security in the 21st century.

A few promising areas for future research are found in: (1) near-real-time 3D reconstruction of a potential build-up of a terror situation; (2) ubiquitous monitoring of airports, harbors, parking lots, and other public areas by the citizenry in general and security personnel in particular; (3) embedded, smart wireless sensor networks streaming such raw data as a building's structural strength, impending structural or ground failures (in a potential earthquake zone) to the Internet; and (4) autonomous robotic agents to take over hazardous safety monitoring work in biological, chemical, and nuclear waste sites. The general research problem area is still wide open for finding elegant solutions to such complex problems as image understanding, 2D/3D modeling, personnel identification, object reconstruction, and behavior descriptions. Finally, much work still needs to be done in fusing together data and techniques from multiple areas of knowledge and expertise to create an intimate awareness of the environment that surrounds an individual, group, or event in order to detect and prevent catastrophic mishaps, be they accidental or intentional.

ACKNOWLEDGMENTS

This work was supported by the DOE University Research Program in Robotics under grant DOE-DE-FG52-2004NA25589 and ONR N00014-03-1-0331.

REFERENCES

ABIDI, B. R. 1995. Automatic sensor placement. In *Proceedings of the SPIE Conference*, 387–398.

ABRAMS, S. AND ALLEN, P. K. 1994. Computing swept volumes for sensor planning tasks. In *Proceedings of the DARPA Image Understanding Workshop*.

ABRAMS, S., ALLEN, P. K., AND TARABANIS, K. A. 1993. Dynamic sensor planning. In *Proceedings of the IEEE Conference on the Robotics and Automation*, 605–610.

AGUILAR, M., FAY, D. A., WAXMAN, M., IRELAND, D. B., RACAMATO, J. P., ROSS, W. D., STREILEIN, W. W., AND BRAUN, M. I. 1998. Fusion of multi-sensor imagery for night vision: Color visualization, target learning and search. In *Proceedings of the SPIE Conference*, 124–135.

AKYILDIZ, I. F., SU, W., SANKARASUBRAMANIAM, Y., AND CAYIRCI, E. 2002. A survey on sensor networks. *IEEE Commun. Mag. 40*, 8, 102–114.

ALAM, M. S., BOGNAR, J. G., HARDIE, R. C., AND YASUDA, B. J. 2000. Infrared registration and high resolution reconstruction using multiple translationally shifted aliased video frames. *IEEE Trans. Instrumentation Measurement 49*, 5, 915–923.

ASHBY, R., AND PRITCHARD, D. A. 2004. Seeing beyond the perimeter: The advanced exterior sensor (AES). In *Proceedings of the 38thAnnual International Carnahan Conference on Security Technology*, 182–188.

AYGUN, R. S. AND ZHANG, A. 2002. Reducing blurring-effect in high resolution mosaic generation. In *Proceedings of the International Conference on Multimedia and Expo*, 537–540.

BABAGUCHI, N., FUJIMOTO, Y., YAMAZAWA, K., AND YOKOYA, N. 2002. A system for visualization and summarization of omnidirectional surveillance video. Tech. Rep. ISIR, Osaka University.

BAKER, S. AND KANADE, T. 2000. Hallucinating faces. In *Proceedings International Conference on Automatic Face and Gesture Recognition*.

BISHOP, C. M., BLAKE, A., AND MARTHI, B. 2003. Superresolution enhancement of video. In *Proceedings of the 9th Workshop on Artificial Intelligence and Statistics*.

BORMAN, S., AND STEVENSON, R. 1998. Super-Resolution from image sequences-A review. In *Proceedings of the Midwest Symposium on Circuits and Systems*, 374–378.

BOSE, B. AND GRIMSON, E. 2004. Learning to use scene context for object classification in surveillance. In *Proceedings of the Joint IEEE International Workshop on Image Understanding*.

BROWN, M. AND LOWE, D. G. 2003. Recognizing panoramas. In *Proceedings of the 9th International Conference on Computer Vision*, 1218–1225.

BUXTON, H. 2003. Learning and understanding dynamic scene activity: A review. *Image Vision Comput. 21*, 125–136.

CAPEL, D. AND ZISSERMAN, A. 1998. Automated mosaicing with super-resolution zoom. In *Proceedings of the International Conference on Computer Vision and Pattern Recognition*, 885–891.

CHANG, S. F., SMITH, J. R., BEIGI, M., AND BENITEZ, A. 1997. Visual information retrieval from large distributed online repositories. *Commun. ACM 40*, 12, 63–71.

CHEN, C. 1998. Image stitching—Comparisons and new techniques. Tech. Rep., Computer Science Department of the University of Auckland, CITR-TR-30, Auckland, New Zealand.

CHEN, S. Y. AND LI, Y. F. 2004. Automatic sensor placement for model-based robot vision. *IEEE Trans. Syst. Man Cybernet. 34*, 1, 393–408.

CHEN, X. AND YANG, J. 2002. Towards monitoring human activities using an omnidirectional camera. In *Proceedings of the 4th IEEE International Conference on Multimodal Interfaces*, 423–428.

CHU, M., HAUSSECKER, H. AND ZHAO, F. 2002. Scalable information-driven sensor querying and routing for ad hoc heterogeneous sensor networks. *Int. J. High Perform. Comput. Appl. 16*, 3, 293–313.

COEN, M. 1998. Design principles for intelligent environments. In *Proceedings of the National Conference on Artificial Intelligence (AAAI)*, 547–554.

COLLINS, R. T., LIPTON, A., FUJIYOSHI, H., AND KANADE, T. 2001. Algorithms for cooperative multisensor surveillance. In *Proceedings of the IEEE International Conference on Computer Vision*, 1456–1477.

COOK, D. J., GMYTRASIEWICZ, P., AND HOLDER, L. B. 1996. Decision-Theoretic cooperative sensor planning. *IEEE Trans. Pattern Anal. Mach. Intell. 18*, 10, 1013–1023.

COWAN, C. K. 1998. Model based synthesis of sensor location. In *Proceedings of the IEEE Conference on Robotics and Automation*, 900–905.

COWAN, C. K. AND KOVESI, P. D. 1998. Automatic sensor placement from vision task requirements. *IEEE Trans. Pattern Anal. Mach. Intell. 10*, 407–416.

CRANE, A. S. 2002. Object recognition with partially labeled examples. Masters thesis, Massachusetts Institute of Technology.

CREBOLDER, J. M., UNRUH, T. D. M., AND MCFADDEN, S. 2003. Search performance using imaging displays with restricted field of view. Tech. Rep., DRDC Toronto TR 2003–007, Defense R&D Canada April..

CRESPI, V., CHUNG, W., AND JORDAN, A. B. 2004. Decentralized sensing and tracking for UAV scheduling. In *Proceedings of the SPIE Conference*.

CROSSBOW TECHNOLOGY. 2008. The gyro guide. http://www.xbow.com/Industry_solutions/gyro_guide.aspx.

CUI, Y., SAMARASEKERA, S., HUANG, Q., AND GREIFFENHAGEN, M. 1998. Indoor monitoring via the collaboration between a peripheral sensor and a foveal sensor. In *Proceedings of the IEEE Workshop on Visual Surveillance*, 2–9.

DAYTON, D., BROWNE, S., SANDVEN, S., GONGLEWSKI, J., GALLEGOS, J., AND SHILKO, M. 2000. Long range laser illuminated imaging. In *Proceedings of the SPIE Conference*.

DE LA ESCALERA, A., ARMINGOL, J. M., AND MATA, M. 2003. Traffic sign recognition and analysis for intelligent vehicles. *Image Vision Comput. 21*, 3, 247–258.

DIEHL, C. P. AND HAMPSHIRE, J. B. 2002. Real-Time object classification and novelty detection for collaborative video surveillance. In *Proceedings of the International Joint Conference on Neural Networks*, 2620–2625.

DRS TECHNOLOGIES. 2008. CBSS—Coastal border surveillance system. http://www.drs.com/products/Coastal_Border.aspx?cat=c3_systems&subcat=Sub_Cat_10.

EIDENBENZ, S. 2002. Approximation algorithms for terrain guarding. *Inf. Process. Lett. 82*, 99–105.

EIDENBERGER, H. 2004. A new perspective on visual information retrieval. In *Proceedings of the SPIE Conference*.

FERREIRA, P. 1994. Interpolation and discrete Papoulis-Gerchberg algorithm. *IEEE Trans. Signal Process. 42*, 10, 2596–2606.

FORSYTH, D. A. AND PONCE, J. 2002. *Computer Vision: A Modern Approach*. Prentice-Hall.

GOODCHILD, M. F. AND LEE, J. 1989. Coverage problems and visibility regions on topographic surfaces. *Annals Oper. Res. 18*, 175–186.

GREER, D. R., FUNG, I., AND SHAPIRO, J. H. 1997. Maximum-Likelihood multiresolution laser radar range imaging. *IEEE Trans. Image Process. 6*, 1, 36–46.

GRINSTEAD, B., KOSCHAN, A., AND ABIDI, M. 2004. Fast digitization of large-scale hazardous facilities. In *Fast Digitization of Large-Scale Hazardous Facilities*, 521–525.

GU, X., MAREFAT, M. M., AND CIARALLO, F. W. 1999. A robust approach for sensor placement in automated vision dimensional inspection. In *Proceedings of the IEEE Conference on Robotics and Automation*, 189–192.

GUNTURK, B. K., BATUR, A. U., ALTUNBASAK, Y. M., HAYES, H. AND MERSEREAU, R. M. 2003. Eigenface-Domain super-resolution for face recognition. *IEEE Trans. Image Process. 12*, 5, 597–606.

GUPTA, A., AND JAIN, R. 1997. Visual information retrieval. *Commun. ACM. 40*, 5, 71–79.

HAGER, G. D. 1990. *Task-Directed Sensor Fusion and Planning—A Computational Approach*. Kluwer Academic.

HALL, D. L. AND LLINAS, J. 1997. An introduction to multisensor fusion. *Proc. IEEE 85*, 1, 6–23.

HAMPAPUR, A., BROWN, L., CONNELL, J., PANKANTI, S., SENIOR, A., AND TIAN, Y. 2003. Smart surveillance: Applications, technologies and implications. In *Proceedings of the IEEE Pacific-Rim Conference on Multimedia*.

HEISELE, B. 2003. Visual object recognition with supervised learning. *IEEE Intell. Syst. 18*, 3, 38–42.

HEO, J., KONG, S., ABIDI, B., AND ABIDI, M. 2004. Fusion of visual and thermal signatures with eyeglass removal for robust face recognition. In *Proceedings of the Joint IEEE International Workshop on Object Tracking and Classification Beyond the Visible Spectrum*, 94–99.

HOOSE, N. 1992. IMPACT: An image analysis tool for motorway analysis and surveillance. *Traffic Eng. Control J.*, 140–147.

HU, W., TAN, T., WANG, L., AND MAYBANK, S. 2003. A survey on visual surveillance of object motion and behaviors. *IEEE Trans. Syst. Man Cybernet. 34*, 3, 334–352.

HUTCHINSON, S. A. AND KAK, A. C. 1989. Planning sensing strategies in a robot work cell with multi-sensor capabilities. *IEEE Trans. Robotics Autom. 5*, 6, 765–783.

IDC. 2006. 2006 U.S. consumer digital imaging survey.

IRANI, M. AND PELEG, S. 1991. Improving resolution by image registration. *CVGIP: Graph. Models and Image Process.* 53, 231–239.

JAVED, O. AND SHAH, M. 2002. Tracking and object classification for automated surveillance. Tech. Rep., Computer Vision Lab, University of Central Florida.

JIA, K. AND GONG, S. 2005. Face super-resolution using multiple occluded images of different resolutions. In *Proceedings of the IEEE Conference on Advanced Video and Signal Based Surveillance*, 614–619.

KANG, J., GAJERA, K., COHEN, I., AND MEDIONI, G. 2004. Detection and tracking of moving objects from overlapping EO and IR sensors. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 123–128.

KANG, S. K., KOSCHAN, A., ZHANG, H. S., PAIK, J. K., ABIDI, B. R., AND ABIDI, M. A. 2002. Hierarchical approach to enhanced active shape model for color video tracking. In *Proceedings of IEEE International Conference on Image Processing*, 888–891.

KANG, S. K., PAIK, J., ABIDI, B. R., SHELTON, D., MITCKES, M., AND ABIDI, M. A. 2001. Gate-to-Gate automated video tracking and location. In *Proceedings of the FAA's 3rd International Aviation Security Technology Symposium*.

KAPLAN, H. AND SCANLON, T. 2001. A thermographer's guide to infrared sensors. Tech. Rep., Honeyhill Technical Company and FLIR Systems.

KASPRZAK, W. 2001. An iconic classification scheme for video-based traffic sensor tasks. In *Computer Analysis of Images and Patterns*. Springer, Berlin.

KASTRINAKI, V., ZERVAKIS, M., AND KALAITZAKIS, K. 2003. A survey of video processing techniques for traffic applications. *Image Video Process. 21*, 359–381.

KEREN, D., PELEG, S., AND BRADA, R. 1988. Image sequence enhancement using sub-pixel displacement. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 742–746.

KIM, K., HWANG, S., SHIN, J., PAIK, J., ABIDI, B., AND ABIDI, M. 2004. Object–Based image restoration for multilayer auto focusing. In *Proceedings of the IS&T/SPIE 16th Annual Symposium on Electronic Imaging Science and Technology*.

Kogut, G. and Trivedi, M. 2002. A wide area tracking system for vision sensor networks. In *Proceedings of 9th World Congress on Intelligent Transport Systems*.

Kong, S. G., Heo, J., Abidi, B. R., Paik, J., and Abidi, M. A. 2005. Recent advances in visual and infrared face recognition—A review. *Comput. Vision Image Understand. 97*, 1, 103–135.

Koschan, A., Kang, S. K., Paik, J. K., Abidi, B., and Abidi, M. 2002. Video object tracking based on extended active shape models with color information. In *Proceedings of 1st European Conference on Color in Graphics, Imaging and Vision*, 126–131.

Krishna, K. M., Hexmoor, H., and Subbarao, P. 2004. Role of autonomy in a distributed sensor network for surveillance. In *Proceedings of the International Conference on Artificial Intellegence IC-AI*, 489–498.

Kuttikkad, S., Meth, R., and Chellappa, R. 1997. Registration and exploitation of multi-pass airborne synthetic aperture radar images. Tech. Rep., Center for Automation Research, University of Maryland.

Lee, J. 1991. Analyses of visibility sites on topographic surfaces. *Int. J. Geograph. Inf. Syst. 5*, 4, 413–429.

Lehel, P., Hemayed, E. E., and Farag, A. A. 1999. Sensor planning for a trinocular active vision system. *IEEE Trans. Comput. Vision Pattern Recogn. 11*, 306–312.

Lertrattanapanich, S. and Bose, N. K. 2002. High resolution image formation from low resolution frames using Delaunay triangulation. *IEEE Trans. Image Process. 11*, 12, 1427–1441

Lin, D., Liu, W., and Tang, X. 2005. Layered local prediction network with dynamic learning for face super-resolution. In *Proceedings of the IEEE International Conference on Image Processing*, 885–888.

Lutz, J. R., Cloud, D., Lareau, A., Bratton, R., and Harris, R. 2004. Control and positioning surveillance sensor technologies for homeland security and homeland defense. In *Proceedings of the SPIE Conference*.

Maas, H.-G. 1998. Photogrammetric techniques for deformation measurements on the reservoir walls. Tech. Rep., Delft University of Technology.

Marcel, B., Briot, M., and Muttieta, R. 1997. Calcul de translation et rotation par la transformation de Fourier. *Traitement du Signal 14*, 2, 135–149.

Margi, C. B. 2003. A survey on networking, sensor processing and system aspects of sensor networks. Tech. Rep., Baskin School of Engineering, University of California at Santa Cruz, February.

Mason, S. 1995. Expert system-based design of close-range photogrammetric networks. *J. Photogrammetry Remote Sensing 50*, 5, 13–24.

Mason, S. and Grun, A. 1995. Automatic sensor placement for accurate dimensional inspection. *Comput. Vision Image Understand. 61*, 3, 454–467.

Matsuyama, T. 2002. Real-Time cooperative multi-target tracking by communicating active vision agents. In *Proceedings of the International Conference on Pattern Recognition*, 14–19.

Mhatre, V., Rosenberg, C., Kofman, D., Mazumdar, R., and Shroff, N. 2004. Design of surveillance sensor grids with a lifetime constraint. In *Proceedings of the European Workshop on Wireless Sensor Networks*, 263–275.

Micheloni, C. and Foresti, G. L. 2003. Fast good features selection for wide area monitoring. In *Proceedings of the Advanced Video and Signal Based Surveillance*, 271–276.

Miller, D. 1999. Landscape visualization using data derived from digital photogrammetry. Tech. Rep., University of California at Santa Barbara.

Miller, J., Kelly, J., Ireland, R., and Byers, J. 1998. Performance of a gimbaled fiber optic coupled laser spotlight and active television. *Optic. Eng. 37*, 10, 2703–2709.

Miura, J., Kanda, T., and Shirai, Y. 2000. An active vision system for real-time traffic sign recognition. In *Proceedings of the IEEE Conference on Intelligent Transportation Systems*, 52–57.

Monson, T., Grantham, J., Childress, S., Sackos, J., Nellums, R., and Lebien, S. 1999. Characterization of scannerless LADAR. In *Proceedings of the SPIE Conference*, 409–420.

Murai, S., Phenekeo, V., Tao, G., Ono, K., and Yasue, S. 1998. Generation of 3D view map using polygon shift method. In *Proceedings of the IAPRS Commission IV Symposium on GIS 32*, 4, 417–421.

Control and Ocean Surveillance Center. 1997. Distributed surveillance sensor network. Annual Rep. to ONR - ONR-3220M.

Nayar, S. K. and Boult, T. E. 1998. Omnidirectional vision systems: 1998 PI report. In *Proceedings of the DARPA Image Understanding Workshop*, 93–100.

Nelson, C. L. 2004. 3-Dimensional video motion detection and assessment. In *Proceedings of the 38th Annual International Carnahan Conference on Security Technology*, 295–302.

Nicolescu, M., Medioni, G., and Lee, M. 2000. Segmentation, tracking and interpretation using panoramic video. In *Proceedings of the IEEE Workshop on Omnidirectional Vision*, 169–174.

Nielsen, M. 2004. *Coverage Analysis*. Masters thesis, Danmarks Tekniske Universitet. August.

OFFICE OF TECHNOLOGY ASSESSMENT. 2003. Airborne platforms and sensors. In *Verification Technologies: Cooperative Aerial Surveillance*.

OLAGUE, G. AND MOHR, R. 2002. Optimal camera placement for accurate reconstruction. *Pattern Recogn. 35*, 927–944.

O'MALLEY, P. D., NECHYBA, M., AND ARROYO, A. 2002. Human activity tracking for wide-area surveillance. In *Proceedings of the Florida Conference on Recent Advances in Robotics*.

O. S. U. P. LAB. 2004. Surface reconstruction and object recognition of urban scenes from multisensor data sets. Lecture Notes GS830, Class 4. May.

PALMER, S. 2002. The semantic Web: An introduction. In *Proceedings of the Website Design and Development Issues for Local Authority Conference*.

PAPASIN, R., GAWDIAK, Y., MALUF, D., LEIDICH, C., AND TRAN, P. 2001. Airport remote tower sensor systems. Tech. Rep. 2001-01-2651, Society of Automotive Engineers, Incorporated.

PARK, S. C., PARK, M. K., AND KANG, M. G. 2003. Super-Resolution image reconstruction: A technical overview. *IEEE Signal Process. Mag. 20*, 3, 19–20.

PARKER, L. AND EMMONS, B. 1997. Cooperative multi-robot observation of multiple moving targets. In *Proceedings of the IEEE Conference on Robotics and Automation*, 2082–2089.

PELEG, S. AND BEN-EZRA, M. 1999. Stereo panorama with a single camera. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 1395–1401.

PELEG, S. AND HERMAN, J. 1997. Panoramic mosaics by manifold projection. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 338–343.

PELEG, S., ROUSSO, B., RAV-ACHA, A., AND ZOMET, A. 2000. Mosaicing on adaptive manifolds. *IEEE Trans. Pattern Anal. Mach. Intell. 22*, 10, 1144–1154.

PHILLIPS, P. J., GROTHER, P., MICHEALS, R. J., BLACKBURN, D. M., TABASSI, E., AND BONE, M. 2002. Face recognition vendor test 2002. Evaluation report.

PONSFORD, A. M. 2004. Detection and tracking of go-east boats using high frequency surface wave radar. In *Proceedings of the SPIE Conference*.

PORIKLI, F. 2004. Multi-Camera surveillance: Object-Based summarization approach. Tech. Rep. TR-2003-145, Mitsubishi Electric Research Laboratory.

POULIN, P., OUIMET, M., AND FRASSON, M.-C. 1998. Interactively modeling with photogrammetry. In *Proceedings of Rendering Techniques*, 93–104.

POWER, C. AND BOULT, T. E. 2000. Evaluation of an omnidirectional vision sensor for teleoperatd target detection and identification. In *Proceedings of the Vehicle Teleoperation Interfaces Workshop*.

PRICE, K. 2007. *Annotated Computer Vision Bibliography.* University of Southern California.

RAY, S. F. 1994. *Applied Photographic Optics*, 2nd ed. Focal Press.

ROUSSO, B., PELEG, S., FINCI, I., AND RAV-ACHA, A. 1998. Universal mosaicing using pipe projection. In *Proceedings of International Conference on Computer Vision*, 945–950.

ROY, S. D., CHAUDHURY, S., AND BANERJEE, S. 2004. Active recognition through next view planning: A survey. *Pattern Recogn. 37*, 429–446.

RUI, Y. AND HUANG, T. S. 2000. Optimizing learning in image retrieval. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 1236–1243.

RUI, Y., HUANG, T. S., ORTEGA, M., AND MEHROTRA, S. 1998. Relevance feedback: a power tool for interactive content-based image retrieval. *IEEE Trans. Circ. Syst. Video Technol. 88*, 2, 1–13.

SAADATSERESHT, M., SAMADZADEGAN, F., AND AZIZI, A. 2004. Accessibility analysis in camera placement network design for vision metrology. In *Proceedings of the FIG Working Week* (May).

SCHOOLEY, M. G. AND THOMPSON, D. 2004. Ethernet-Based integrated surveillance system for homeland security and homeland defense. In *Proceedings of the SPIE Conference*.

SCOTT, W. R., ROTH, G., AND RIVEST, J.-F. 2003. View planning for automated three-dimensional object reconstruction and inspection. *ACM Comput. Surv. 35*, 1, 64–96.

SCOTTI, G., MARCENARO, L., COELHO, C., SELVAGGI, F., AND REGAZZONI, C. S. 2005. Dual camera intelligent sensor for high definition 360 degrees surveillance. *Vision, Image, Signal Process. 152*, 2, 250–257.

SEASTRAND, M. J. 2004. Maritime microwave radar and electro-optical data fusion for homeland security. In *Proceedings of the SPIE Conference*.

SHADBOLT, N., BERNERS-LEE, T., AND HALL, W. 2006. The semantic Web revisited. *IEEE Intell. Syst. 21*, 3, 96–101.

SMEULDERS, A. W. M., WORRING, M., SANTINI, S., GUPTA, A., AND JAIN, R. 2000. Content-Based image retrieval at the end of the early years. *IEEE Trans. Pattern Anal. Mach. Intell. 22*, 12, 1349–1380.

SMITH, D. AND SINGH, S. 2006. Approaches to multisensor data fusion in target tracking: a survey. *IEEE Trans. Knowl. Data Eng. 18*, 12, 1696–1710.

SMITH, P., SHAH, M., AND LOBO, N. DA V. 2004. Integrating and employing multiple levels of zoom for activity recognition. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 928–935.

SOATTO, S., FREZZA, R., AND PERONA, P. 1994. Motion estimation on the essential manifold. In *Proceedings of the European Conference on Computer Vision*, 61–72.

SOERGEL, U., THOENNESSEN, U., AND STILLA, U. 2001. Utilization of LIDAR DEM for SAR image analysis in dense urban areas. *Int. Archives Photogrammetry Remote Sensing. 34*, 3B, 255–258.

STEINVALL, O., SOEDERMAN, U., AHLBERG, S., SANDBERG, M., LETALICK, D., AND JUNGERT, E. 1999. Gated viewing for target detection and target recognition. In *Proceedings of the SPIE Conference*, 432–448.

STORCH, M. T. 2004. Cell-Tower deployment of counter-sniper sensors. In *Proceedings of the SPIE Conference*.

SZELISKI, R. 1996. Video mosaics for virtual environments. *IEEE Comput. Graph. Appl. 16*, 2, 22–30.

TAN, J., AND XI, N. 2004. Peer-to-Peer model for the area coverage and cooperative control of mobile sensor networks. In *Proceedings of the SPIE Conference*, vol. 5403.

TARABANIS, K. A., ALLEN, P. K., AND TSAI, R. Y. 1995. A survey of sensor planning in computer vision. *IEEE Trans. Robotics Autom. 11*, 1, 86–104.

TARABANIS, K. A., TSAI, R. Y., AND ALLEN, P. K. 1995. The MVP sensor planning system for robotic vision tasks. *IEEE Trans. Robotics Autom. 11*, 1, 72–85.

TRIVEDI, M. M., PRATI, A., AND KOGUT, G. 2002. Distributed interactive video arrays for event based analysis of incidents. In *Proceedings of the Intelligent Transportation Systems*, 950–956.

TSIA, R. Y. AND HUANG, T. S. 1984. Multiframe image restoration and registration. In *Advances in Computer Vision and Image Processing*, JAI Press, 317–339.

TURK, M. 1998. Moving from GUIs to PUIs. In *Proceedings of the Symposium on Intelligent Information Media*.

UKITA, N. AND MATSUYAMA, T. 2002. Real-Time cooperative multi-target tracking by communicating active vision agents. In *Proceedings of the 16th International Conference on Pattern Recognition*, 14–19.

UR, H. AND GROSS, D. 1992. Improved resolution from sub-pixel shifted pictures. *CVGIP: Graph. Models Image Process. 54*, 2, 181–186.

VANDEWALLE, P., SUSSTRUNK, S., AND VETTERLI, M. 2005. A frequency domain approach to registration of aliased images with applications to super-resolution. *EURASIP J. Appl. Signal Process.* 233–233.

VANIJJA, V. AND HORIGUCHI, S. 2001. 360° interactive video scenes with multi-directional moving capability. In *Proceedings of the 11th International Conference on Artificial Reality and Telexistence*, 209–212.

VON DER FECHT, A. AND ROTHE, H. 1998. Comparison of imaging laser radars based on range gating under different weather conditions. In *Proceedings of the SPIE Conference*, vol. 3380, 292–302.

WAGNER, R., WAAGEN, D., AND CASSABAUM, M. 2004. Image super-resolution for improved automatic target recognition. In *Proceedings of the SPIE Defense and Security Symposium*.

WANG, F. AND QI, H. 2004. A cognitive information processing framework for distributed sensor networks. In *Proceedings of the SPIE Conference*, vol. 5403.

WANG, J., YAN, W.-Q., KANKANHALLI, M. S., JAIN, R., AND REINDERS, M. J. T. 2003. Adaptive monitoring for video surveillance. In *Proceedings of IEEE Pacific-Rim Conference on Multimedia*.

YAO, Y., ABIDI, B., AND ABIDI, M. 2006. Fusion of omnidirectional and PTZ cameras for accurate cooperative tracking. In *Proceedings of the IEEE International Conference on Advanced Video and Signal Based Surveillance*.

YAO, Y., ABIDI, B., KALKA, N. D., SCHIMID, N., AND ABIDI, M. 2007. Super-Resolution for high magnification face images. In *Proceedings of the SPIE Defense and Security Symposium, Biometric Technology for Human Identification IV*.

YASTIKLI, N. AND JACOBSEN, K. 2003. Automatic digital elevation model generation, problems and restrictions in urban areas. Tech. Rep., Yildiz Technical University.

YE, Y. AND TSOTSOS, J. K. 1996. Sensor planning for 3D object search. In *Proceedings of the International Symposium on Intelligent Robotic Systems*.

YI, S., HARALICK, R. M., AND SHAPIRO, L. G. 1992. Automatic sensor and light source positioning for machine vision. In *Physics-Based Vision: Theory and Practice*. Jones and Bartlett, 394–398.

ZLATANOVA, S. AND TEMPFLI, K. 2000. Modeling for 3D GIS: Spatial analysis and visualization through the Web. In *Proceedings of the International Association for Primate Refuges and Sanctuaries Symposium*, vol. XXXIII.

ZHAO, F., SHIN, J., AND REICH, J. 2002. Information-Driven dynamic sensor collaboration for tracking applications. *IEEE Signal Process. Mag. 19*, 2, 61–72.

Received December 2004; revised December 2007; accepted February 2008