Multimodal Surveillance: Sensors, Algorithms and Systems, Zhigang Zhu and Thomas S. Huang (Co-Editors), Artech House Publisher, to appear in June 2007

Chapter 1. Multimodal Surveillance: an Introduction

Zhigang Zhu (CCNY), Thomas Huang (UIUC)

In the Merriam-Webster Online Dictionary, the word “surveillance” literally means "watching over" (in French) [1]. In Wikipedia, an online free encyclopedia, surveillance is defined as “the process of monitoring the behavior of people, objects or processes within systems for conformity to expected or desired norms in trusted systems for security or social control” [2].

Surveillance began as the art of watching over the activities of persons or groups from a position of higher authority. Surveillance has been an intrinsic part of human history. Sun Tzu's *The Art of War*, written in Chinese about 2,500 years ago, discusses in one whole chapter (among 13 chapters) how spies should be used against enemies [3]:

“Thus, what enables the wise sovereign and the good general to strike and conquer, and achieve things beyond the reach of ordinary men, is ***foreknowledge***. Now this foreknowledge cannot be elicited from spirits; it cannot be obtained inductively from experience, nor by any deductive calculation. Knowledge of the enemy's dispositions ***can only be obtained from other men***.” (The Art of War, Chapter XIII: The Use of Spies)

This tells us the “foreknowledge” cannot be obtained from reasoning, experience or calculation; it can only be obtained from other men (i.e., spies, human “sensors”). Then, five different classes of “other men” were discussed in great details. In fact, nearly a millennium earlier, in *Numbers*, one of the five law books supposedly written by Moses between 1440 and 1400 B.C., twelve men, each from one of the twelve Israeli ancestral tribes, were sent out to “explore” the land of Canaan before Israelites went into the “promised land” [4]:

“When Moses sent them to explore Canaan, he said, ‘Go up through the Negev and on into the hill country. See what the ***land*** is like and whether the ***people*** who live there are strong or weak, few or many. What ***kind of land*** do they live in? Is it good or bad? What ***kind of towns*** do they live in? Are they unwalled or fortified? How is the **soil**? Is it fertile or poor? Are there ***trees*** on it or not? Do your best to bring back some of the ***fruit*** of the land.’ (It was the season for the first ripe grapes.)” (Numbers 13:17-20)

Interestingly, this gives us a picture of multi-sensory and multi-modal explorations for an important surveillance mission. It was multi-sensory (‘twelve men”). It was multimodal (not only exploring “people” and their attributes, but also “land”, facilities, soil, vegetations, etc). It was distributed (“go up through” and “on into”). Finally, Moses also needed to resolve differences or even conflicts in the information gathered (e.g., “we worked like grasshoppers to them” versus “we will swallow them up” in the paragraphs following the verses quoted above).

Of course, it is modern electronic and computer technologies that have pushed surveillance towards the science end of "watching over". Closed circuit TV (CCTV) - where the picture is viewed or recorded, but not broadcasted- initially developed as a means of security for banks. Today it has developed to the point where it is simple and inexpensive enough to be used in home security systems, and for everyday surveillance. However, although video surveillance is the most popular form of surveillance, there are all forms of observation or monitoring, not just visual observation. In fact, the word "surveillance" has historically been used to describe observation from a distance by means of electronic equipment or other technological means, for example: eavesdropping, telephone tapping, directional microphones, communications interception, covert listening devices or 'bugs', subminiature cameras, closed-circuit television, GPS tracking, electronic tagging, reconnaissance aircraft, and etc. Therefore, surveillance in nature should include multiple modalities.

The greatest impact of computer-enabled surveillance is due to involvement of a large number of organizations in surveillance operations. The state and security services still have the most powerful surveillance systems, because they are enabled under the law. But today's levels of state surveillance have increased in the era of anti-terrorism, and using computers they are now able to draw together many different information sources with multimodal "sensors" to produce profiles of persons or groups in society. Many large corporations now use various form of 'passive' surveillance. This is primarily a means of monitoring the activities of staff and for controlling public relations. But some large corporations such as supermarkets and financial firms actively use various forms of surveillance to monitor the activities of customers who may impact their operations. In military, Intelligence, Surveillance, Target Acquisition, and Reconnaissance (ISTAR) [5] links several battlefield functions together to assist a combat force in employing its sensors and managing the information they gather. Information is collected on the battlefield through systematic observation by deployed soldiers and a variety of electronic sensors. Surveillance, Target Acquisition and Reconnaissance are methods of obtaining this information. The information is then passed to Intelligence personnel for analysis, and then to the commander and his staff for the formulation of battle plans.

In recent years, we witness a rapid growing of research and development of surveillance and biometrics using multimodal sensors, including video, audio, thermal, vibration, and various kinds of other sensors, in both civilian and military applications. This edited book is a collection of sample works contributed by leading researchers in the emerging field of multimodal surveillance. We roughly group the works into three equally important parts:

(1) multimodal sensor and novel sensing approaches;
(2) multimodal data fusion algorithms; and
(3) multimodal system issues.

A multimodal surveillance system usually consists of all the three components. Sensors are important since they are the front-ends of a surveillance system. Algorithms are important since they are the cores to enable the system to work. Other system-level issues are also important for building a real multimodal system. In the book we also include works for various tasks of surveillance applications. Most of the chapters address *people and activity issues* in tasks such as human and vehicle tracking, human identification via biometrics, speech detection and recognition, etc. A few chapters are also dealing with *environment modeling and monitoring issues* for surveillance, mainly in Chapter 15 on multimodal 3D scene modeling, and some in chapter 4 on target selections for remote hearing, and in Chapter 11 and Chapter 17 on sentient computing and sentient spaces.

The criterion for including a chapter into the book is that it addresses some aspects of multimodal surveillance. We also consider the distributions of the works from different parts of the world, and the balance of the contributors from academia, government and industry. In the following we will give a brief summary of each chapter, in the context of each of the three parts.

1.1 Mutlimodal Sensor and Sensing Approaches

In the first part, we include four chapters (Chapters 2-5) on multimodal sensors and sensing approaches. The sensors discussed are either multimodal sensors in nature (e.g. in Chapters 2 and 3), or are novel sensors that are components of multimodal surveillance systems (e.g., in Chapters 4 and 5). Some of them are high-end sensors for high fidelity and/or long range surveillance, whereas others target low-cost solutions.

ARL Multi-Modal Sensor (Chapter 2)

In Chapter 2, Houser and Zong describe the ARL Multi-Modal Sensor, which is a research tool for target signature collection, algorithm validation, and emplacement studies at the U.S. Army Research Laboratory (ARL). ARL has a significant program involving the development of Unattended Ground Sensors (UGS) that address a variety of military and government missions. ARL's program involves practically every aspect of sensor development including devices, detection and fusion algorithms, communications, and command and control. One element of the ARL UGS program involves the development of low cost sensing techniques for the urban environment and one embodiment of this effort is the Multi-Modal Sensor (MMS). The program objectives of this effort were to develop a ***networked personnel detection sensor*** with the following major criteria: low cost in volume production, support MOUT (Military Operations in Urban Terrain) operations, and employ non-imaging sensor diversity techniques. The MMS sensor was an early prototype intended to demonstrate that low cost sensing techniques were suitable for the urban environment and a viable alternative to higher cost and fidelity sensors for some applications. The MMS is used today as a demonstration system and a test bed for many facets of urban sensing. This chapter will describe many aspects of the MMS design including: hardware, software, and communications. The detection algorithms will also be described including collection of target signatures and validation of algorithm performance. Finally, MMS usage in a ***force protection*** application will be described including issues encountered when integrating into a larger system.

Visible-Thermal Biometric Systems (Chapter 3)

In Chapter 3. Socolinsky discusses the smart deployment of visible-thermal biometric surveillance systems at Equinox Corporation. The combined use of visible and thermal infrared imagery has proven valuable for a wide range of surveillance tasks, including facial recognition. However, the added value must be weighed against the increased cost and complexity, as well as the unique challenges that arise from the use of a bi-modal system. The primary question is when and why we should specify a fused imaging solution. Whenever such a solution is appropriate, then we must strive to provide the simplest system meeting the required performance. This will insure the most reliable operation, shortest time to deployment and lowest operational cost. In this chapter, the author explores some of the issues surrounding design and deployment of fused visible-thermal systems for ***biometric surveillance***, including appropriate choice of sensor wavelengths for given scenarios, optical designs best suited to particular tasks, and cost/performance tradeoffs. The problems of cross-modality matching and availability of legacy data is considered in the context of ***face recognition***. While human facial recognition is used as the focus of discussion, much of the chapter is applicable to other forms of surveillance.

Laser Doppler Vibrometry for Remote Hearing (Chapter 4)

In Chapter 4, Zhu and his research team at the City College of New York discuss a novel laser Doppler vibrometry sensor for ***remote hearing*** in a multimodal surveillance system for human signature detection. The system consists of three types of sensors: an infrared (IR) camera, a pan/tilt/zoom (PTZ) electro-optical (EO) color camera and the laser Doppler vibrometer (LDV). In particular, the LDV is explored as a new non-contact remote voice detector. In their research, they have found that voice energy vibrates most objects (targets) in the surrounding environment of the speaker, and the vibrations can be detected by an LDV at a large distance. Since signals captured by the LDV are very noisy, algorithms have been designed with Gaussian bandpass filtering, Wiener filtering and adaptive volume scaling to enhance the LDV voice signals. The enhanced voice signals are intelligible from LDV signals returns via targets without retro-reflective finishes at short or medium distances (< 100 m). By using retro-reflective tapes, the distance could be as far as 300 meters. However, the authors point out that manual operations to search and focus the laser beam on a target (with both proper vibration and reflection) are very difficult at medium and large distances. Therefore, IR/EO imaging for ***target selection and localization*** is also discussed. Future work remains in LDV sensor improvement, automatic LDV targeting and intelligent refocusing for long range LDV listening.

Audio-Assisted Cameras and Acoustic Doppler Sensors (Chapter 5)

In Chapter 5, novel low-cost sensor and data systems are presented by Smaragdis, Raj and Kalgaonkar at Mitsubishi Electric Research Labs in Cambridge. Specifically, they describe two technologies: ***audio-assisted cameras***, and acoustic Doppler sensors for ***gait recognition***. They have made various arguments on why audio analysis can be a useful complement to video cameras and have shown how Doppler sensing can be applied for gait recognition at a very low cost. Although the two subjects were treated separately they are by no means meant to be kept as such. Future plans involve the convergence of these two projects that will allow the enhancement of audio and camera systems by Doppler sensing, and help improve the Doppler estimates by multi-modal inputs.

1.2 Multimodal Fusion Algorithms

In the second part, we include six chapters (Chapters 6-11) on multimodal integration of various sensing modalities, using various algorithms. The first two chapters (Chapters 6 and 7) in this part discuss multimodal audio-visual integration for two different tasks – automatic speech recognition, and human detection and tracking. Chapters 8, 9 and 10 discuss various approaches of multimodal biometrics, by combining face recognition with three other different types of sources – ear, palm and gait, respectively. Finally in Chapter 11, a multi-sensory and multi-modal approach is presented for sentient computing with a large-scale indoor surveillance system.

Automatic Audio-Visual Speech Recognition (Chapter 6)

In Chapter 6, Chu (IBM T.J. Watson Research Center) and Huang (UIUC) present a novel method for automatic audio-visual speech recognition. This chapter considers the fundamental problem of multimodal fusion in the context of pattern recognition tasks in surveillance and human-computer interactions. The authors propose a novel sensory fusion method based on the ***coupled hidden Markov Models (CHMM)*** for audio-visual speech modeling. The CHMM framework allows the fusion of two temporally coupled information sources to take place as an integral part of the statistical modeling process. An important advantage of the CHMM-based fusion method lies in its ability to model asynchronies between the audio and visual channels. They describe two approaches to carry out inference and learning in CHMM. The first is an exact algorithm derived by extending the forward-backward procedure used in HMM inference. The second method relies on the model transformation strategy that maps the state space of a CHMM onto the state space of a classic HMM, and therefore facilitates the development of sophisticated audio-visual speech recognition systems using existing infrastructures. Audio-visual speech recognition experiments on two different corpora validate the proposed the fusion approach. In both evaluations, it is clearly demonstrated that the CHMM-based intermediate integration scheme can utilize the information in the visual channel more effectively than the early integration methods in noisy conditions. The audio-visual systems can achieve higher recognition accuracy than audio-only systems even when matched acoustic conditions are ensured.

Multimodal Tracking via Cameras & Microphone Arrays (Chapter 7)

Many applications such as interactive multimedia, videoconferencing, and surveillance require the ability to track the 3-D motion of the subjects. ***Particle filters*** represent an attractive solution for the tracking problem because they do not require solution of the inverse problem of obtaining the state from the measurements and because the tracking can naturally integrate multiple modalities. In Chapter 7, Zotkin, Raykar, Duraiswami and Davis at the University of Maryland present a framework for multimodal tracking using multiple cameras and multiple microphone arrays. In order to calibrate the resulting distributed multi-sensor system, the authors propose a method to automatically determine the 3-D positions of all microphones in the system using at least five loudspeakers. The method does not require knowledge of the loudspeaker positions but assumes that for each loudspeaker there exists a microphone very close to it. They derive the ***Maximum Likelihood estimator***, which reduces to the solution of the non-linear least squares problem. A closed-form approximate solution that can be used as an initial guess is derived. They also derive an approximate expression for the estimator covariance using the implicit function theorem and Taylor series expansion. Using the estimator covariance matrix, the authors analyze the performance of the estimator with respect to the positions of the loudspeakers; in particular, they show that the loudspeakers should be as far away from each other, and the microphones should lie within the convex hull formed by the loudspeakers. They verify the correctness and robustness of the multimodal tracker and of the self-calibration algorithm both with Monte-Carlo simulations and on real data from three experimental setups. The authors also present practical details of system implementation.

Multimodal Biometrics: Face + Ear (Chapter 8)

Multimodal biometrics is one of the important functionalities in a multimodal surveillance system. In Chapter 8, a multimodal biometrics system involving the human ear is presented by Middendorff and Bowyer at the University of Notre Dame. The authors note that recent years have seen substantial and increasing interest in exploring the use of the human ear as a source for biometrics. In any biometric scenario where one might acquire images of the face, it images of the ear could also potentially be acquired. Interestingly, reports in the literature suggest that, using the same ***PCA-style recognition algorithm*** and face and ear images of similar quality of the same subjects, the recognition rate achieved using the ear as a biometric is similar to that achieved using the face. The combination of the ear and the face then has the potential to achieve higher accuracy. Also, a biometric system that could analyze both ear and face could still produce a result if either one is not able to be imaged for some reason. Therefore, there are both accuracy and coverage advantages to be gained in a multi-modal ear + face biometric system. Researchers have also explored using both 3D and 2D data for ear biometrics. Results of several studies suggest that 3D shape of the ear allows greater recognition accuracy than 2D appearance. This is likely due to pose variation between the gallery and probe images being more readily handled in case of 3D shape matching. However, 3D sensors are currently both more expensive than normal cameras and also more restrictive in terms of conditions for image acquisition. For this reason, the authors explore the degradation of 2D "***eigen-ear***" performance due to pose change between gallery and probe, and consider approaches to dealing with this problem.

Multimodal Biometrics: Face + Palmprint (Chapter 9)

In Chapter 9, researchers in the Center for Biometrics and Security Research, Chinese Academy of Sciences, present a new multimodal biometric identification system by fusing face and palmprint features to improve the identification performance. Effective classifiers based on the so-called ***ordinal*** features are first constructed from images of faces and palmprints, respectively. Then, several strategies including ***Sum, Product, Max and Min rules*** and ***linear discriminant analysis (LDA)*** are employed for fusing them on a data set that consists of face and palmprint data of 378 subjects, 20 pairs for each. Experimental results have shown the effectiveness of the proposed system.

Multimodal Biometrics: Face + Gait (Chapter 10)

Recognition of humans from arbitrary viewpoints is an important requirement for different applications such as intelligent environments, surveillance and access control. For optimal performance, the system must use as many cues as possible from appropriate vantage points and fuse them in meaningful ways. In Chapter 10, Chellappa and his research team at the University of Maryland discuss issues of human identification by the fusion of face and gait cues from a monocular video. The authors use a view invariant gait recognition algorithm for gait recognition and a face recognition algorithm using particle filters. They employ decision fusion to combine the results of gait and face recognition algorithms, and consider two fusion scenarios: ***hierarchical*** and ***holistic***. The first employs the gait recognition algorithm when the person is far away from the camera and passes the top few candidates to the face recognition algorithm. The second approach involves combining the similarity scores obtained individually from the face and gait recognition algorithms. Simple rules like the SUM, MIN and PRODUCT are used for combining the scores. The results of fusion experiments are demonstrated on the NIST database which has outdoor gait and face data of thirty subjects.

Multi-Sensory Sentient Computing (Chapter 11)

Finally in Chapter 11, Town in the University of Cambridge Computer Laboratory presents an approach to multi-sensory and multi-modal (more than 2) fusion in which computer vision information obtained from calibrated cameras is integrated with a large-scale indoor surveillance system known as "SPIRIT". The SPIRIT system employs an ultrasonic location infrastructure to track people and devices in an office building and model their state in order to provide a ***sentient computing environment*** for context aware applications. The vision techniques used include background and object appearance modeling, face detection, segmentation, and tracking modules. Integration is achieved at the system level through the metaphor of ***shared perceptions***, in the sense that the different modalities are guided by and provide updates to a shared world model. This model incorporates aspects of both the static (e.g. positions of office walls and doors) and the dynamic (e.g. location and appearance of devices and people) environment. Fusion and inference are performed by ***Bayesian networks*** that model the probabilistic dependencies and reliabilities of different sources of information over time. It is shown that the fusion process significantly enhances the capabilities and robustness of both sensory modalities, thus enabling the system to maintain a richer and more accurate world model.

1.3 Multimodal System Issues

In the third part, we include six chapters on various system issues in building and/or supporting mutlimodal surveillance systems. Those issues include: information representations, system architectures /frameworks /workbenches, usability issues, real-time performance, 24/7 operations, automatic environment modeling, system evaluations, and distributed infrastructure (middleware).

E-Chronicling System for Mobile Human Surveillance (Chapter 12)

The first work in this part is about a multimodal personal mobile surveillance system. Rapid advances in mobile computing devices and sensor technologies are enabling the capture of unprecedented volumes of data by individuals involved in field operations in a variety of applications. As capture becomes ever more rich and pervasive, the biggest challenge lies in the developments of ***information processing and representation tools*** that maximize the utility of the captured multi-sensory data. The right tools hold the promise of converting captured data into actionable intelligence resulting in improved memory, enhanced situational understanding, and more efficient execution of operations. These tools need to be at least as rich and diverse as the sensors used for capture, and need to be unified within an effective ***system architecture***. In Chapter 12, an initial attempt to an end-to-end e-chronicling system for mobile human surveillance is presented by Pingali and his colleagues at the IBM T.J. Watson Research Center. The system combines several emerging sensor technologies, state of the art analytic engines, and multi-dimensional navigation tools, into an end-to-end electronic chronicling solution.

Usability Issues in Distributed Multimodal Surveillance (Chapter 13)

Surveillance systems inherently have the human tightly coupled in the system loop. In most cases, the human operator is the decision maker who will act upon (or not act upon) the feedback of the system. In the aspect of usability, video surveillance systems are different from other software applications in that they are time critical and emphasize on accuracy of users' responses. Decision often needs to be made accurately in real time. Furthermore, what is of interest to a particular surveillance system user can vary greatly, and the security forces using the system are not, in general, advanced computer users. Therefore the usability issue becomes crucial to a multi-modal surveillance system. In Chapter 13, such systems issues in distributed multi-modal surveillance is discussed by Yu (ObjectVideo Inc.) and Boult (University of Colorado at Colorado, Spring). The authors argue that viable commercial multi-modal surveillance systems need to be ***reliable***, ***robust,*** and they must able to ***work at night*** (maybe the most critical time). They must handle small and non-distinctive targets that are as far away as possible. Like other commercial applications, end users of the systems must be able to operate them in a proper way. In this chapter, the authors focus on three significant inherent limitations of current surveillance systems: the effective accuracy at relevant distances, the ability to define and visualize the events on a large scale, and the usability of the system.

Multimodal Workbench for Automatic Surveillance (Chapter 14)

In Chapter 14, a multimodal workbench for automatic surveillance applications is presented by Datcu, Yang and Rothkrantz at Delft University of Technology in Netherlands. The framework (i.e., the workbench) is designed to facilitate the communication between different processing components. The authors argue that in modern surveillance applications, satisfactory performance will not be achieved by a single algorithm, but rather by a combination of interconnected algorithms. The proposed framework specification only emphasizes the presence and role of its processing components. In other words, it only considers the description of the processing components along with their interconnections, not the implementation details. This framework is centered on the ***shared memory paradigm***, which allows loosely coupled asynchronous communication between multiple processing components, both in time and space. The shared memory in the current design of the framework takes the form of eXtended Markup Language (XML) data spaces. This suggests a more human-centric paradigm to store, retrieve and process the data. Because the framework implementation itself relies on the philosophy of shared XML data spaces, special attention is given on how to integrate the two underlying technologies, namely, the processing components and XML data management. An example is given of a surveillance application built on top of the framework. The application uses several cameras and microphones to detect unusual behavior in train compartments. The authors note that although the framework has been designed for an automatic surveillance oriented application, it can be adopted as basis for any kind of complex multimodal system involving many components and heavy data exchange. The specification fully complies with the requirements of data manipulation in a multi data producer/consumer context where the availability of data is time-dependent and some connections might be temporarily interrupted.

Multimodal Scene Modeling for Urban Surveillance (Chapter 15)

Three-dimensional models of urban environments are useful in a variety of applications such as urban planning, training and simulation for disaster scenarios, virtual heritage conservation, and urban surveillance. For surveillance applications, the models provide a lot of useful information about background, context and occlusions to facilitate moving target detection and tracking. In Chapter 15, Zakhor at the University of California at Berkeley presents a fast approach to automated generation of textured 3D city models with both high details at ground level, and complete coverage for bird's-eye view, using multimodal air and ground sensing approaches. A close-range facade model is acquired at the ground level by driving a vehicle equipped with ***laser scanners*** and a ***digital camera*** under normal traffic conditions on public roads; a far-range Digital Surface Map (DSM), containing complementary roof and terrain shape, is created from ***airborne laser scans***, then triangulated, and finally texture mapped with ***aerial imagery***. The facade models are registered with respect to the DSM by using Monte-Carlo-Localization, and then merged with the DSM by removing redundant parts and filling gaps. The developed algorithms are evaluated on a data set acquired in downtown Berkeley and elsewhere.

Usage Scenarios & Evaluation of Multimodal Biometrics (Chapter 16)

To fully assess the utility of multimodal biometric techniques in real-world systems, it is useful to consider usage scenarios and applications in which multimodal techniques are implemented. In Chapter 16, Thieme presents results from an evaluation conducted by International Biometric Group (IBG) and funded by the National Institute of Justice. The evaluation compared the accuracy of various multimodal fusion and normalization techniques based on data generated through commercial ***fingerprint***, ***face*** recognition, and ***iris*** recognition systems. This evaluation attempted to situate results in the context of typical biometric applications and usage scenarios.

Middleware in Multimodal Sentient Environments (Chapter 17)

In Chapter 17, researchers at the University of California at Irvine present SATware: a Stream Acquisition and Transformation (SAT) middleware they are developing to analyze, query, and transform ***multimodal sensor data streams*** to facilitate flexible development of ***sentient environments***. A sentient space possesses capabilities to perceive and analyze situation based on data acquired from disparate sources. A multimodal stream processing model in SATware and its elementary architectural building blocks are discussed. These include a distributed runtime system that permits injection, execution, and interconnection of stream processing operators, a declarative language for composition of such operators, an operator deployment module that optimizes deployment of stream processing operators, the concept of ***virtual sensors*** that encapsulates stream processing topologies to create semantic sensing abstractions, and an infrastructure directory for storing the availability of resources. The authors describe how this basic architecture provides a suitable foundation for addressing the challenges in building middleware for customizable sentient spaces. SATware is implemented in the context of Responsphere—a pervasive computing, communication, and sensing ***infrastructure*** deployed at UC Irvine that serves as a unique testbed for research on situation monitoring and awareness in emergency response applications. SATware provides a powerful application development environment in which users (i.e., application builders) can focus on the specifics of the application without having to deal with the technical peculiarities of accessing a large number of diverse sensors via different protocols.

1.4 Concluding Remarks

Multimodal sensing, data processing, and system integration are still in a fast growing stage of research and development. Much work needs to be done to ensure successful real-world, large-scale surveillance applications. The collection in this book is just samples of some representative works in the field and therefore by no means to be complete. However, we hope this collection will stimulate more interests in the research and development of mutlimodal surveillance techniques, including the further developments of mutlimodal sensors, multimodal data fusion algorithms, and mutlimodal surveillance systems.

References

[1]. Merriam-Webster Online Dictionary, *Surveillance*, http://www.m-w.com/dictionary/surveillance
[2]. Wikipedia, the free encyclopedia, *Surveillance*, http://en.wikipedia.org/wiki/Surveillance
[3]. Sun Tzu, *The Art of War*. Translated from the Chinese by Lionel Giles, M.A. (1910), http://www.chinapage.com/sunzi-e.html
[4]. Moses, *Numbers*, New International Version, Chapter 13, http://www.biblegateway.com/passage/?search=Numbers%2013;&version=31;
[5]. Wikipedia, *ISTAR*, http://en.wikipedia.org/wiki/ISTAR