Systems issues in distributed multi-modal surveillance

Li Yu
ObjectVideo Inc.
11600 Sunrise Valley Dr.
Reston, VA 20191
liyu@objectvideo.com

Terrance E. Boult
University of Colorado at Colorado Springs
1420 Austin Bluffs Parkway
Colorado Springs, CO 80933-7150
tboult@vast.uccs.edu

Abstract

To be viable commercial multi-modal surveillance systems, the systems need to be reliable, robust and must be 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 paper, we 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.

1. User Interfaces

Visualization and usability are crucial to the success of video surveillance systems. Video surveillance systems are not only measured by statistic results of its vision module (e.g. detection and false alarm rate), but also the degree of user satisfaction of the whole system. Surveillance systems inherently have the end users tightly coupled in the system loop (Figure 1). In most cases, the human is the decision maker who will act upon (or not act upon) the feedback of the system. As shown in Figure 1, a typical surveillance system takes frames from video cameras as input, processes frames with vision module, and generates system feedbacks (in the forms of alerts, email notification, and etc.). Generally human operators are involved in two types of tasks: define rules for vision module and monitor the system's feedbacks. Well-designed graphical user interfaces (GUI) can help users understand the strength and limitation of the system, and thus operate the system in a proper way. On the other hand, poor-designed GUI can compromise the performance of the whole system if end users do not know how to interpret what is displayed on the monitor and make poor decisions.

Our goal is to address the usability issue of a video surveillance system by presenting a paradigm to integrate the GUI design aspect of a surveillance system with underlying vision engines [3]. A well-designed GUI cannot only improve the usability but also the performance of the vision modules. The paradigm is called Understanding Images of Graphical User Interfaces (UI-GUI, pronounced as ōō ’ē g ōō ’ ē). The basic idea of UI-GUI is that a video surveillance system visualizes the states of the vision engine onto the GUI. Based on what they see on the screen, users define rules. The visual rules, as part of the GUI, are extracted and used to generate rule logics. At run time, detection and/or tracking results of the vision module will be matched up against the rule logics. Alerts are triggered if rules are broken.

Research has been done on how to evaluate the usability of a software system. The usability of a general graphical user interface (GUI) is categorized into five areas: Overall Reaction, Learning, Screen, Terminology, and System Capability [2]. To evaluate a system, users are asked to perform tasks with the system and make quantitative evaluation in the five categories.

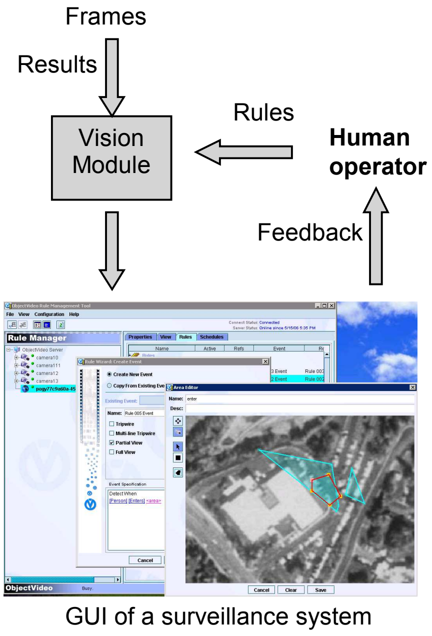


Figure 1: A generic architecture of a video surveillance system. Human operators are tightly coupled with the system loop.

2. System Issues in Large Scale Video Surveillance

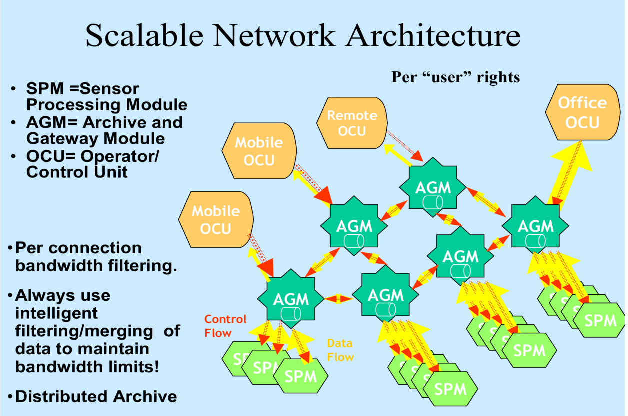
Intelligent video surveillance is a systems level problem with 5 major components:

1. Low-level detection/processing algorithms
2. Higher-level algorithms for combining data and selecting particular events of interest
3. User-Interface
4. Sensor/Hardware/Computation architecture
5. Software/Communication architecture

The second part of the paper takes a more general systems view, reviewing what was necessary to develop solutions that support commercial intelligent camera networks deployed with hundreds of sensors. The two most fundamental issues for the distributed surveillance systems are tied to the resolution needed for the tasks and the communication. The section reviews the key system components in effective “distributed video surveillance,” then discusses the major open issues, including hardware-accelerated algorithms needed for increasing resolution while reducing power, and the issues of mobile surveillance.

The section reviews issues in tradeoffs between target size, field of view and response time and cost. It then discusses active visual surveillance, which uses data from computer controlled Pan/Tilt/Zoom (PTZ) units combined with state of the art video detection and tracking to, in a cost effective manner, provide active assessment of potential targets. This active assessment allows an increase in the number of pixels on target and provides a secondary viewpoint for data fusion, while still allowing coverage of a very large surveillance area. This active approach and multi-sensor fusion, not a new concept, was developed as part of the DARPA Video Surveillance and Monitoring (VSAM) program in the late 90's (Figure 2). While we have continued to expand upon it since that time, there has been only limited academic research and until 2003, no commercial video surveillance, that provided these important abilities.

The final system issue discussed is that of data/network communication. While some systems assume analog cable feeds back to a central location, this is quite costly and does not scale extremely well. We discuss the communication issues for a distributed system and the extensions we provided to the original DARPA VSAM protocol to support adaptive bandwidth management to allow hundreds of cameras on wireless links [1].



Scalable Network Architecture

- SPM =Sensor Processing Module
- AGM= Archive and Gateway Module
- OCU= Operator/ Control Unit

- Per connection bandwidth filtering.
- Always use intelligent filtering/merging of data to maintain bandwidth limits!
- Distributed Archive

Figure 2: Scalable extended VSAM Architecture with distributed adaptive bandwidth control

References

[1] A Lipton, T. Boult, and Y. Lee, "Video surveillance and monitoring communication specification document 98-2.2," tech. rep., CMU, Sept. 1998.

[2] J. Chin, V. Diehl, K. Norman, "Development of an Instrument Measuring User Satisfaction of the Human-Computer Interface," CHI’88 Conference Proceedings. NY: ACM Press, 213-218, 1988.

[3] L. Yu, T. E. Boult, "Understanding Images of Graphical User Interfaces: A New Approach to Activity Recognition for Visual Surveillance," Second IEEE Workshop on Event Mining, in conjunction with IEEE CVPR04.