Design, Implementation and Performance Analysis of Pervasive Surveillance Networks

Amit Goradia[†], **Zhiwei Cen**[‡], **Clayton Haffner**[‡], **Ning Xi**[†] and **Matt Mutka**[‡]
[†]Department of Electrical and Computer Engineering
[‡]Department of Computer Science and Engineering
Michigan State University
East Lansing, Michigan, 48824, USA

Abstract

Pervasive surveillance implies the continuous tracking of multiple targets as they move about the monitored region. The tasks to be performed by a surveillance system are expressed as the following requirements:(1) Automatically track the identified targets over the region being monitored: (2) Provide concise feedback and video data of a tracked target to multiple operators. The active sensors needed to track the target keep changing due to target motion. Hence in order to provide concise and relevant information to a human operator to assist in decision making, the video feedback provided to the operator needs to be switched to the sensors currently involved in the tracking task. Another important aspect of surveillance systems is the ability of track multiple targets simultaneously using sensors with motion capability. Current feature (point) based visual surveillance and tracking techniques generally employed do not provide an adequate framework to express the surveillance task of tracking multiple targets simultaneously using a single sensor. This paper presents a mutational analysis approach for shape based control to model a multi-target surveillance scenario. A surveillance testbed has been designed based on these requirements and the proposed algorithms and subsystems are implemented on it and a performance analysis of proposed methods have been provided.

Introduction

Networked surveillance systems provide an extended perception and distributed reasoning capability in monitored environments through the use of multiple networked sensors. The individual sensor nodes can have multiple sensing modalities such as cameras, infrared detector arrays, laser range finders, omnidirectional acoustic sensors, etc. Locomotion and active sensing greatly increase the range and sensing capability of the individual sensor nodes. Multiple nodes facilitate simultaneous multi-view observation over a wide area and can aid in reconstruction of 3D information about the tracked targets. A pervasive surveillance network (PSN) is comprised of a collection of active sensor nodes equipped with visual sensing, processing, communication and motion capabilities.

The surveillance network must provide a timely and concise view of the relevant activities within the environment being monitored to the human operator. Providing multiple video feedback streams often causes loss of attention span of the operator and makes it hard to keep track of the various activities over the various cameras. Hence only video streams from relevant sensors should be presented to the operator on a per activity basis. This would involve automatically switching the active video stream presented to the operator. This paper presents an analysis of the realtime video transport protocols (MJPEG, H.263) implemented with specific relevance to the video surveillance task.

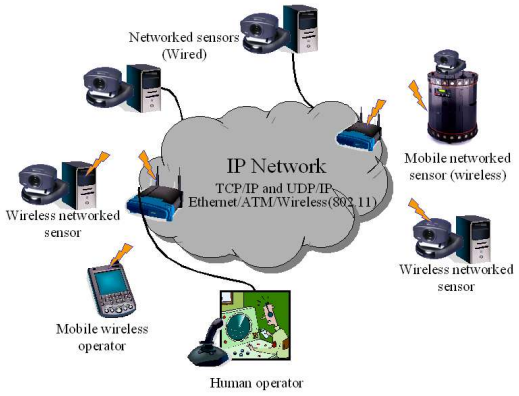

Figure 1: Networked Surveillance.

The identified targets can be tracked using visual tracking algorithms, such as visual servo control (Hutchinson, Hager, & Corke 1996) or gaze control (Brown 1990) , which mainly involve feature (point) based tracking and fail to describe the basic task of maintaining the target in the sensor's active field of view effectively and succinctly. These approaches cannot address the problem of ensuring the coverage of multiple targets using a single sensor. In order to overcome the above mentioned problems with automated control approaches found in literature we propose the image based Hausdorff tracking method which tries to ensure that the multiple targets specified for the tracking task do not leave the active FOV of the sensor and the size of the target sets are maintained at a discernable resolution.

Major contributions of this paper include proposing an automated multiple target tracking framework using active cameras. A surveillance system with the perceived goals is implemented and the performance of design alternatives for the switched video feedback subsystem are measured and analyzed.

Networked Surveillance

Networked surveillance systems have received much attention from the research community due to their many pervasive applications (Regazzoni, Ramesh, & Foresti 2001). Our implementation of a visual surveillance system consists of multiple heterogenous sensor nodes, with video cameras

mounted on them, connected to each other over and IP based communication network. The nodes have the processing, communication and limited motion capabilities. The general architecture of the surveillance system is shown in figure 1. The tasks to be performed by a surveillance system can be effectively expressed as the following requirements:

1. Identify multiple moving targets based on predefined models
2. Automatically track the identified targets over the region being monitored
3. Provide concise feedback and video data of a tracked target to multiple operators

Video Feedback

Video feedback is the essential component of the surveillance system. The operator needs the video to make decisions about the tracking task. Automatic image analysis and video understanding tools (R.T.Collins *et al.* 2001) can be facilitated to activate alarms or logs for certain surveillance tasks.

Since multiple cameras are deployed to track the specified targets, the feedback video streams required for monitoring the target might be more than one at the same time and will be changing from time to time. Providing multiple un-necessary (un-related to the task) video feedback streams often causes loss of attention span of the operator and makes it hard to keep track of the various activities over the cameras. Hence only video streams from relevant sensors should be presented to the operator on a per activity basis. This is done through automatic or manual switching of the camera streams that are presented to the operator.

In the implementation section of this paper we compared different video encoding and real time transport protocols such as MJPEG (Berc *et al.* 1998) and H.263 (ITU-H.263 1996) transported over RTP transport protocol (Schulzrinne *et al.* 1997). Their performance under certain situations are measured and the advantages and disadvantages of different protocols are analyzed.

Automated Surveillance

In order to perform automated surveillance there are two major subtasks: target detection and target tracking. A target perception and video understanding module is responsible for detecting and classifying the various targets in the active field of view (FOV) of the sensor and performing temporal consolidation of the detected targets over multiple frames of detection. Moving target detection and classification is known to be a difficult research problem and has been the focus of many recent research efforts (Toyoma *et al.* 1999). Many approaches such as active background substraction (T.Matsuyama & N.Ukita 2002) and temporal differentiation have been suggested for detecting and classifying various types of moving targets including single humans and human groups to vehicles and wildlife (R.T.Collins *et al.* 2001).

When a target is recognized in the active FOV of a sensor, it can be tracked using image based tracking methods like visual servoing and gaze control (R.T.Collins *et al.* 2001) (T.Matsuyama & N.Ukita 2002). However, these approaches only try to maintain an image feature (point) at the center of the screen and the algorithm used are very sensitive to feature detection and do not express the objectives of the task adequately. Another significant disadvantage of these techniques is that they can describe the tracking task for only one target at a time. However, in a wide area surveillance scenario a sensor may be tasked with maintaining visibility of multiple targets at a time.

In order to solve the active target tracking problem, we propose to use a mutational analysis approach (Aubin 1993). Multiple target coverage can be readily expressed in a set based topological framework using shape analysis and shape functions ((Cea 1981) (Sokolowski & Zolesio 1991)). Thus, the variables to be taken into account are no longer vectors of parameters but the geometric shapes (domains) themselves. Unfortunately, due to the lack of a vectorial structure of the space, classical differential calculus cannot be used to describe the dynamics and evolution of such domains. Mutational analysis endows a general metric space with a net of "directions" in order to extend the concept of differential equations to such geometric domains. Using mutational equations, we can describe the dynamics (change in shape) of the sensor field of view (FOV) and target domains and further derive feedback control mechanisms to complete the specified task.

The surveillance task can be expressed, using shape functions (Cea 1981), as the minimization of a Hausdorff distance based metric or the size of the target etc. The shape function essentially represents the error between the desired and actual shapes and reducing it to zero will accomplish the task. The remainder of this section presents the method of Hausdorff tracking using mutational equations for performing the surveillance task.

Hausdorff Tracking

Shape or a geometric domain can be defined as the set $K \in \mathcal{K}(E),\ E \subset \mathbb{R}^n$ where $\mathcal{K}(E)$ represents the space of all nonempty, compact subsets of E. The target and the camera coverage can be readily expressed as shapes. Mutational equations can then be used to express the change (deformation) in the coverage and target sets based on the motion of the sensor. Shape analysis (Cea 1981) can be used to address problems involving geometric domains or shapes. Shape functions, which are set defined maps from $J(K) : \mathcal{K}(E) \mapsto \mathbb{R}$, can be used to provide a "measure" of acceptability and optimality of the shape K. For example we can use a shape function to see if a reference set $\hat{K}$ is contained within a current set K. In order to accomplish the task defined using shape functions, we need to derive a feedback map $\mathcal{U} : \mathcal{K}(E) \mapsto U$, where $u = \mathcal{U}(K(t))$ is the input to the sensor, which will reduce the shape function to zero. The convergence of the shape function can be analyzed using the shape Lyapunov theorem (Doyen 1994). The convergence to zero of the task function would imply task accomplishment.

Target, Coverage Sets and Shape Functions The target blob is represented as the set $\hat{K}$ of pixels comprising it and the sensor coverage set is represented as rectangle centered at the image center, K, as shown in figure 2. The task requirements of maintaining the target within the active FOV of the sensor with an adequate resolution can be mathematically expressed as a shape function having the form:

$$J(\hat{K}) = \int_{\hat{K}} f(q)\, dq \tag{1}$$

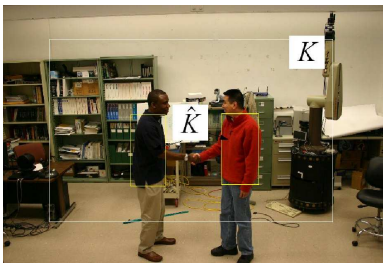

Figure 2: Targets and coverage set for image based Hausdorff tracking.

where, $q \in \hat{K}$ and $f(q)$ is a function of the resolution of the target image and the directed Hausdorff demi-distance distance $d_K(q) = \{\|q - p\| \; \|p \in K, q \in \hat{K}\}$ of the target $\hat{K}$ from the coverage set K.

Dynamics Model Using Mutational Equations The deformation of the target set w.r.t. the motion of the camera can be represented using a mutational equation and can be modeled using optic flow equations.

Assuming that the projective geometry of the camera is modeled by the perspective projection model, a point $P = [x, y, z]^T$, whose coordinates are expressed with respect to the camera coordinate frame, will project onto the image plane with coordinates $q = [q_x, q_y]^T$ as:

$$\begin{bmatrix} q_x \\ q_y \end{bmatrix} = \frac{\lambda}{z} \begin{bmatrix} x \\ y \end{bmatrix} \tag{2}$$

where λ is the focal length of the camera lens (Hutchinson, Hager, & Corke 1996). Using the perspective projection model of the camera, the velocity of a point in the image frame with respect to the motion of the camera frame (Hutchinson, Hager, & Corke 1996) can be expressed. This is called the image Jacobian by (Hutchinson, Hager, & Corke 1996) and is expressed as:

$$\begin{bmatrix} \dot{q}_x \\ \dot{q}_y \end{bmatrix} = \varphi_c(q) = B_c(q) \begin{bmatrix} u \\ \dot{\lambda} \end{bmatrix} = B_c(q) u_c \tag{3}$$

$$B_c(q) = \begin{bmatrix} -\frac{\lambda}{z} & 0 & \frac{q_x}{z} & \frac{q_x q_y}{\lambda} & -\frac{(\lambda^2+q_x^2)}{\lambda} & q_y & \frac{q_x}{\lambda} \\ 0 & -\frac{\lambda}{z} & \frac{q_y}{z} & \frac{\lambda^2+q_y^2}{\lambda} & -\frac{q_x q_y}{\lambda} & -q_x & \frac{q_y}{\lambda} \end{bmatrix}$$

where, $u = [v_x,\ v_y,\ v_z,\ \omega_x,\ \omega_y,\ \omega_z]^T$ is the velocity screw of the camera motion and $\dot{\lambda}$ is the rate of change of the focal length.

Using above equation 3 the mutational equation (Aubin 1993; Goradia *et al.* 2005) of the target set can be written as:

$$\begin{aligned} \dot{q} &= \varphi(q) = \varphi(q) \\ \mathring{\hat{K}} &\ni \varphi(\hat{K}) \end{aligned} \tag{4}$$

Feedback Map u The problem now is to find a feedback map u_c such that the shape function J is reduced to zero. For this purpose we need to find the shape directional derivative $\mathring{J}(\hat{K})(\varphi)$ of $J(\hat{K})$ in the direction of the mutation $\varphi(\hat{K})$. From (Aubin 1993) and (Sokolowski & Zolesio 1991), the directional derivative of the shape function having the form of equation 1 can be written as:

$$\mathring{J}(\hat{K})(\varphi) = \int_{\hat{K}} \text{div}(f(q)\varphi(q))\, dq \tag{5}$$

Assuming a relatively flat object, i.e., the z coordinate of all the points on the target are approximately the same we can derive an expression for $\mathring{J}(\hat{K})(\varphi)$ by substituting equations 1 and 4 into 5 as:

$$\mathring{J}(\hat{K})(\varphi) \leqslant \begin{bmatrix} \frac{1}{z_t} C_1(q) & C_2(q) \end{bmatrix} u_c \tag{6}$$

where,$u_c = [v_x, v_y, v_z, \dot{\lambda}, \omega_x, \omega_y, \omega_z]^T$ and z_t is an estimated minimum bound on the target z position and $z_t > z$ will guarantee the inequality in 6.

Using the shape Lyapunov theorem (Doyen 1994), we can find the assumptions on input u_c such that the shape function $J(\hat{K})$ tends to zero as:

$$\begin{bmatrix} \frac{1}{z_t} C_1(q) & C_2(q) \end{bmatrix} u_c \leqslant -\alpha J(\hat{K}) \tag{7}$$

The feedback map u_c which is an input to the camera module can be calculated from the above equation 9 using the notion of a generalized pseudoinverse $C^{\#}(q)$ of the matrix $C = \begin{bmatrix} \frac{1}{z_t} C_1(q) & C_2(q) \end{bmatrix}$ as:

$$u_c = C^{\#}(\alpha J(\hat{K})) \tag{8}$$

It should be noted that the estimate z_t of the target distance only affects the gain of the control and not its validity. Further it is important to note that the gain distribution between the various redundant control channels depends on the selection of the null space vector when calculating the generalized pseudoinverse $C^{\#}$ of matrix C.

Surveillance System Requirements and Implementation

We have built a pervasive surveillance network testbed to demonstrate the integration of multiple active sensors with active target tracking algorithms to perform a coherent pervasive surveillance task of tracking multiple targets as they move across the monitored landscape. The testbed consists of multiple active cameras attached to processing, communication units mounted on pan-tilt drives or robots for moving the cameras. The surveillance testbed developed has the functionality of an end-to-end, multicamera monitoring system which allows a single or multiple human operator(s) to monitor activities in the region of surveillance.

The architecture of the implemented systems is shown in figure 1. It consists of multiple active camera sensors interconnected using an IP network which consists of wired ethernet as well as wireless links. There are multiple clients which can pass queries to the network regarding the respective targets they want to track. Visual feedback is provided to the clients based on the queries they have requested. The remainder of this section provides the details of the implementation of the surveillance testbed.

System Hardware

The sensor node setup consists of three Sony EVI-D30 active PTZ (pan-tilt-zoom) cameras as shown in figure 3. The cameras were connected to Pentium 4 2.4 GHz computers which had PCI based video capture hardware cards attached to them. The PTZ drives are controlled through serial port communications. The various computers were connected to each other using wired ethernet and wireless 802.11g connection over and IP network. The individual sensor nodes were provided with publicly addressable IP address and hence could

Figure 3: Sony EVI-D30 cameras with PTZ drives.

be accessed from the Internet. The human interface clients could be connected to the surveillance network through a direct wired or wireless connection or through the Internet.

Video Subsystem

The video subsystem is designed to support video feedback to multiple clients over an IP network. The video service is designed to support various types of live video stream feedback form the sensors to the individual clients such as MJPEG and H.263. Various resolutions such as CIF, QCIF and 4CIF are supported for the live video streams. In order to transmit real-time video over a network, it is necessary to convert the output of the encoder into a packet sequence i.e., packetize it. Traditional TCP and UDP services are not sufficient for real-time applications and applications oriented protocols such as the Real-time Transport Protocol RTP provide an alternate solution for managing the essential tradeoffs for quality and bandwidth. RTP provides end-to-end delivery services such as payload type identification, sequence numbering and time stamping. Characteristics and implementation details of the various video subsystems implemented are summarized in the following discussion.

MJPEG MJPEG stands for Motion-JPEG and is a technique which simply performs a JPEG compression on each video frame before transmission. Unlike MPEG and H.263 codecs, MJPEG does not support temporal compression but only supports spatial compression.

The main advantages to using this approach is that JPEG compression can be implemented relatively easily in hardware and it supports a wide variety of resolutions. This implies that a wide variety of hardware can be supported incase of a network with heterogenous video capture hardware. Further it uses no inter-frame compression which results in low latency in the video transmission system. However, the major disadvantage for using MJPEG technology is its inefficient use of bandwidth. Due to the lack of inter-frame (temporal) compression, MJPEG streams require a high bandwidth of the order of 2 Mbits/s for a 30 fps NTSC resolution stream. Though at lower frame rates and lower resolutions MJPEG can be used effectively, its use cannot be justified in low bandwidth applications such as wireless sensor networks.

H.263 H.263 is video coding standard by the International Telecommunications Union (ITU). It was designed for data rates as low as 20 Kbits/s and is based on the ITU H.261 standard. It supports 5 resolutions (CIF, QCIF, sub-QCIF, 4CIF and 16CIF, where CIF is standard 352×288 pixels resolution). It uses both temporal and spatial compression and provides for advanced coding options such as unrestricted motion vectors, advanced prediction and arithmetic coding (instead of variable length coding) for improvement of video quality at the expense of video codec complexity. It allows for fixed bit rate coding for transmission over a low bandwidth network as well as variable bit rate coding for preserving a constant image quality and frame rate for storage and transmission over high bandwidth networks.

Automated Tracking

The sensor outputs are processed on the local computer for target detection and identification. The targets to be tracked using the automated tracking system were multiple humans moving around in the monitored region. For ease of target detection and identification, the human targets were wearing solid color clothing and *CMVision* was used for color analysis and blob detection and merging (Bruce, Balch, & Veloso 2000). The acquired 640×480 pixel images were quantized on a 128×96 grid for reduced computation load. The identified targets were represented using their bounding boxes which were quantized on the 128×96 grid. The coverage set is also quantized on the grid and the distances of the grid points to the target set are pre-computed and stored.

The automated tracking task was defined as maintaining the visibility of multiple moving targets in a region with adequate resolution. The shape function used to track the targets is:

$$J(\hat{K}) = \sum_{i=1}^{N} J_{FOV}(\hat{K}_i) + J_{Amin}(\hat{K}_i) + J_{Amax}(\hat{K}_i) \quad (9)$$

$$J_{FOV}(\hat{K}_i) = \int_{\hat{K}_i} d_K^2(p)\, dq$$

$$J_{Amin}(\hat{K}_i) = \max(\int_{\hat{K}_i} dq - AREA\_MIN_i,\ 0)$$

$$J_{Amax}(\hat{K}_i) = \min(AREA\_MAX_i - \int_{\hat{K}_i} dq,\ 0)$$

where, N is the number of targets, q is a point on the target set $\hat{K}$ and $AREA\_MAX_i$ and $AREA\_MIN_i$ denote the maximum and minimum admissible areas of the target set $\hat{K}_i$ for maintaining adequate resolution. Note that the shape function $J(\hat{K})$ is zero only when set of targets $\bigcup_{i=1}^{N} \hat{K}_i$ is completely covered by the sensor coverage set K and when the area of each target set $\hat{K}$ is within the limits $(AREA\_MIN, AREA\_MAX)$ specified for that target. Otherwise $J(\hat{K})$ is a non-zero positive value.

Based on the value of the shape function $J(\hat{K})$, the velocity input vector u to the camera motion units (PTZ drives or robot) is calculated and applied at the rate of image acquisition i.e., 25 frames per second (fps)(Goradia *et al.* 2005).

Architecture of Sensor Node

Figure 4 the general architecture of a sensor node. The target perception module is responsible for detecting and classifying the various targets in the active field of view (FOV) of the sensor and performing temporal consolidation of the detected targets over multiple frames of detection. The video transmission consists of a video server which compresses and transmits the video information from the sensor node to clients requesting the video information. Compression and transmission of the video stream is accomplished using MJPEG or H.263 bit stream mounted over the RTP/UDP/IP transport protocol.

The individual sensor nodes maintain information regarding the observations of their neighboring nodes and broadcast (within their locality) their own observations. Based on

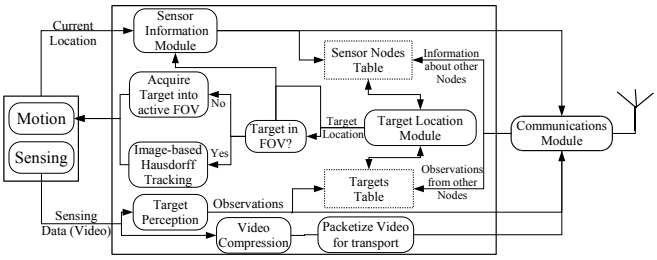

Figure 4: Architecture of Sensor Node.

the combined observations, each node develops a list of targets being actively tracked and the status of its peer nodes and stores this information in the targets table and the sensor nodes table, respectively. In the targets table, the native as well as observed characteristics of the target objects, observed by the respective sensors are stored. The targets table also stores information indicating the node that sensed these characteristics. Nodes also store peer information, such as location, active FOV and total capable FOV of the peer. When targets is recognized in the active FOV of a sensor, they can be tracked using image based Hausdorff tracking.

Performance Analysis of the Implemented System

Video system

The real-time performance video subsystem implemented was measured and analyzed for use in the surveillance network scenario. The parameters which affect the performance of a video transport scheme in the context of a switched surveillance scenario are time taken to deliver the video stream to the operator, size and quality of the image, the rate of video frame update (frame rate) and the initialization time for the client to receive one complete image frame from the server. We compared the performance of using H.263 and MJPEG as the compression scheme for the visual surveillance tasks.

Video size and image quality The H.263 standard is very limited in its capability for selection of the frame size which is limited to 5 standard resolutions namely, CIF, QCIF, SIF, 4CIF and 16CIF. This may be a limitation when integrating multiple types of camera sensors. On the other hand MJPEG allows for almost all resolutions and can be used to transmit very high definition images when required. The viewing quality of the transmitted images can be largely regarded as same as both the schemes use DCT (discrete cosine transform) and quantization for compression. However, it should be noted that using MJPEG the complete image is updated at once while for H.263 the image is updated in parts and may cause a visual degradation of the perceived image quality.

Video bitrate and frame rate Due to the inter-frame (temporal) compression implementation, the video bitrate generated per frame for H.263 is lower than compared to MJPEG. This makes H.263 more suitable for video communications over a restricted bandwidth network. The frame rates and bitrates for the two schemes generated for various quantization Q's (Berc *et al.* 1998) are tabulated in table 1.

Table 1: MJPEG and H.263 frame rate v/s bitrate

MJPEG				H.263			
Q	FPS	Bitrate(KBPS)		Q	FPS	Bitrate(KBPS)	
		panning	static			panning	static
30	25	1700	1700	10	25	1300	48
	20	1300	1300		20	1000	38
	10	668	668		10	700	20
70	25	3000	3000	1	25	1900	175
	20	2500	2500		20	1300	156
	10	1200	1200		10	1000	106

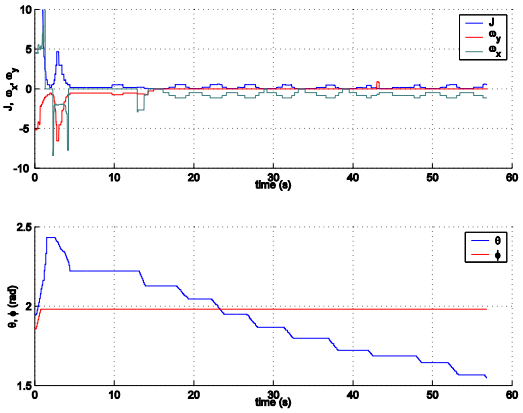

Figure 5: Image based Hausdorff tracking.

Server switching for camera handoff The initialization time of both H.263 and MJPEG is also measured. The initialization time is defined as the time from when the server starts broadcasting to when the client receives the first image frame. In both cases, the initialization is around 10ms. However for H.263, since inter-frame coding is used, there will be long delay when a client joins a session where server is already broadcasting to other clients. The effect is termed as "late entry". This is because for inter-frame encoding, the server is only required to transmit those master blocks that have changed through consecutive changes. Thus the client has to wait until all the master blocks to be transmitted once to be able to see the whole scene. The delay can be large especially when the scene is relatively static. The H.263 standard recommends all the master blocks be transmitted once per 132 frames. In our experiment, when we set the frame rate to be 1 fps, H.263 has to wait for approximately 32 seconds to receive the the complete image.

However, it should be noted that for the H.263 coding scheme, all the information needed to initialize the decoder is stored in an intra coded 'I' reference frame, which is transmitted periodically. One method to combat this late entry problem would be transmit an 'I' frame every time a new client joins the session and can be implemented using the RTCP (control commands) part of the RTP protocol. This may lead to higher bandwidth consumption which again could be acceptable in surveillance applications.

Automated Target Tracking with Active Camera

The surveillance task is to maintain the multiple targets in the active FOV of the sensor. The targets were two humans moving around and interacting with each other. Assumptions on the input $u = [\omega_x, \omega_y]^T$ to the camera system were derived

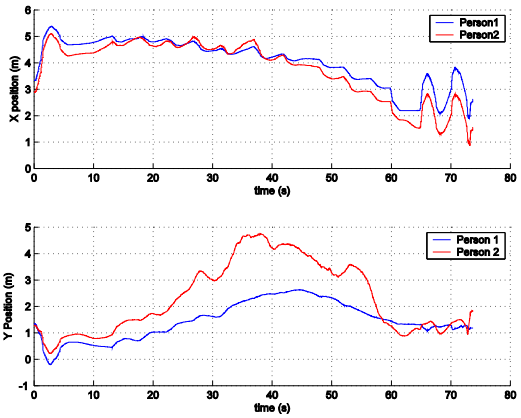

Figure 6: Image based Hausdorff Tracking: Estimated position of targets.

using equation 8, where ω_x, ω_y are the pan and tilt velocities. At, $t = 0$, the targets are just in the active FOV of the sensor and task criterion is not satisfied. The camera then moves to reduce the shape function J to zero so the targets are now covered. The targets then randomly move around the room and the camera tries to maintain both targets continuously in the active FOV.

Figure 5 depicts the J and the input velocities $u = [\omega_x, \omega_y]^T$ applied to the camera. Notice the initial large value of the shape function J, which is quickly reduced to zero. Figure 6 depicts the position X, Z estimated of the two targets. We see that despite the seemingly random motion of the two targets, the camera always tries to keep both of them in the active FOV. Further, the energy efficiency of the proposed method is demonstrated by the relatively infrequent input applied to the camera only when one of the objects escapes the active FOV.

Discussions and Conclusions

This paper presents the design and implementation of a pervasive surveillance system for multi target tracking. In order to implement the system we developed a realtime multiple target tracking framework using mutational analysis and a switched video transport interface which enables the concise interaction of a human operator with the network.

The major advantage of the Hausdorff tracking algorithm is that it can be used to succinctly describe a tracking task involving multiple targets. The capability of defining the tracking task involving multiple targets also allows flexibility for the system to be overloaded and un-encumbered by the restriction placed on most surveillance systems for number of targets being actively tracked being less than the number of sensors involved (T.Matsuyama & N.Ukita 2002).

The video subsystem is implemented with MJPEG and H.263 mounted over RTP transport protocol and a comparative analysis for these two schemes is presented. The MJPEG system can transmit video frames of various sizes and hence has the advantage of being able to handle heterogenous hardware for video capture. Further it has a low coding and initialization latency and allows for a direct hardware implementation which is advantageous for using lower processing power on the sensor node. It does not suffer from the late entry problem during switching. However, it requires a consistently high bandwidth which may limit its application on wireless or low bandwidth communication channels.

For H.263 the encoding schemes requires low bandwidth consumption for relatively static scenes. Thus high frame rate can be achieved. This is quite important for continuous and responsive real-time monitoring. The drawbacks of H.263 is that it can handle video frames of certain specified sizes only and that both initialization time and switching time are higher than MJPEG. It also suffers for the late entry problem which can be detrimental for systems involving video stream switching and nodes linked over un-reliable channels. However, the late entry problem can be solved at the expense of transmitting the complete intra frame encoded image when requested using the RTCP protocol.

References

Aubin, J.-P. 1993. Mutational equations in metric spaces. *Set-Valued Analysis* 1:3–46.

Berc, L.; Fenner, W.; Frederick, R.; McCanne, S.; and Stewart, P. 1998. Rtp payload format for jpeg-compressed video. *RFC2435*.

Brown, C. 1990. Gaze control with interactions and delays. *IEEE Trans. on Systems Man and Cybernetics* 20(1).

Bruce, J.; Balch, T.; and Veloso, M. 2000. Fase and inexpensive color image segmentation for interactive robots. In *IROS*.

Cea, J. 1981. Problems of shape optimal design. *Optimization of Distributed Parameter Structures vol. I and II* 1005–1087.

Doyen, L. 1994. Shape laypunov functions and stabilization of reachable tubes of control problems. *Journal of Mathematical Analysis and Applications* 184:222–228.

Goradia, A.; Xi, N.; Cen, Z.; and Mutka, M. 2005. Modeling and design of mobile surveillance networks using a mutational analysis approach. In *International Conference on Intelligent Robots and Systems*.

Hutchinson, S.; Hager, G.; and Corke, P. 1996. A tutorial on visual servo control. *IEEE Transactions on Robotics and Automation* 12(5):651–670.

ITU-H.263. 1996. Video coding for low bit rate communication. *ITU-T Recommendation H.263*.

Regazzoni, C.; Ramesh, V.; and Foresti, G. E. 2001. Special issue on third generation surveillance systems. *Proc. of the IEEE* 89.

R.T.Collins; A.J.Lipton; H.Fujiyoshi; and T.Kanade. 2001. Algorithms for cooperative multisensor surveillance. *Proceedings of the IEEE* 89:1456–1477.

Schulzrinne, H.; Casner, S.; Frederick, R.; and Jacobson, V. 1997. Rtp: A transport protocol for real-time applications. *RFC1889*.

Sokolowski, J., and Zolesio, J.-P. 1991. *Introduction to Shape Optimization: Shape Sensitivity Analysis*. Computational Mathematics. Springer-Verlag.

T.Matsuyama, and N.Ukita. 2002. Real-time multitarget tracking by a cooperative distributed vision system. *Proceedings of the IEEE* 90(7):1136–1150.

Toyoma, K.; Krumm, J.; Brumitt, B.; and Meyers, B. 1999. Wallflower: Principles and practice of background maintainence. In *International Conference on Computer Vision*, 255–261.