A Design Methodology for Selection and Placement of Sensors in Multimedia Surveillance Systems

Siva Ram G.S.V.S., K.R. Ramakrishnan
Indian Institute of Science
Bangalore, India.
sivaram@ece.iisc.ernet.in,
krr@ee.iisc.ernet.in

P.K. Atrey, V.K.Singh, M.S.Kankanhalli
National University of Singapore,
Singapore.
{pradeepk,vivekkum,mohan}
@comp.nus.edu.sg

ABSTRACT

This paper addresses the problem of how to select the optimal number of sensors and how to determine their placement in a given monitored area for multimedia surveillance systems. We propose to solve this problem by obtaining a novel performance metric in terms of a probability measure for accomplishing the task as a function of set of sensors and their placement. This measure is then used to find the optimal set. The same measure can be used to analyze the degradation in system's performance with respect to the failure of various sensors. We also build a surveillance system using the optimal set of sensors obtained based on the proposed design methodology. Experimental results show the effectiveness of the proposed design methodology in selecting the optimal set of sensors and their placement.

Categories and Subject Descriptors

H.5.1 [**Multimedia Information Systems**]; I.6.4 [**Simulation and Modeling**]: Model Validation and Analysis

General Terms

Design, Security

Keywords

Sensor selection and placement, Performance metric, Fault tolerance

1. INTRODUCTION

Most of the multimedia surveillance systems nowadays utilize multiple types of sensors which have different capabilities and which are of different costs. The design of such systems plays an important role in achieving the required performance. The proper selection and placement of sensors is important because it provides the required performance at minimal cost.

VSSN'06, October 27, 2006, Santa Barbara, California, USA.

In this paper, we consider the following problem. Given a set of sensors to be employed in an environment (a convex region) and also given a surveillance task, we propose a novel design methodology which, in order to accomplish the task with a specified performance, determines - 1) Optimal number of sensors, 2) Their optimal placement and 3) System failure behavior. One can think of this design methodology as a black box which takes inputs like geometry of the convex surveyed area, types of sensors, surveillance task and the desired performance; and outputs the optimal set of sensors and their placement.

The core idea is to obtain a performance metric for accomplishing the given surveillance task as a function of set of sensors and their placement in a given convex surveyed area and then use this measure to find the optimal set of sensors and their placement. Such a performance measure allows the system designer to analyze the degradation in system's performance when some of the sensors fail. Deriving such a performance metric is challenging and requires the modeling of the effect of the interplay of the individual sensors when placed in a particular configuration. While the performance metric is highly task dependent, the methods that we propose in this paper are general and thus useful for all designers of surveillance systems. We consider a surveillance task of capturing the frontal information of a symmetric object in a convex surveyed region. This task is chosen as it is a common task across many systems and the performance metric that we derive for this task can be applied to other tasks such as object detection and tracking etc. Also, examples of symmetric objects like human and animal faces, cars etc. are frequently encountered in surveillance scenarios.

In this paper, we describe our design methodology by considering two types of sensors which are PTZ (Pan-Tilt-Zoom) infrared cameras and active motion sensors. We also build a surveillance system consisting of these two types of sensors for object tracking. Experimental results confirm that optimal sensor placement based on design maximizes the system performance.

Thus our main contribution in this paper is to propose a design methodology for multimedia surveillance systems that helps a system designer in optimally selecting and placing the sensors in order to accomplish a given task with a specified performance. The proposed design methodology is 'directionally aware' (i.e. realizes that only images obtained in a certain direction may be useful) and can easily scale to multiple PTZ cameras as well as motion sensors. To the best of our knowledge, this is the first time that a design strategy has been proposed for building such heterogeneous

surveillance systems. Our interactions with the industry also indicate that an ad hoc methodology is usually employed for designing such systems.

The remainder of this paper is organized as follows. Section 2 presents the related work. In section 3, we describe the proposed design methodology for selecting the optimal set of sensors and their placement. In section 4, we discuss the system implementation and then present the experimentation results in section 5. Finally, in section 6, we conclude the paper with a discussion on future work.

2. RELATED WORK

In the past, optimal sensor selection problem has been studied in the context of discrete-event systems and failure diagnosis. Oshman [7] proposes to select sensors at each epoch based on the information gain along the state space direction. Debouk et al [2] proposed to identify instances where it is possible to explicitly determine optimal strategies for the markovian decision problem. [5] is based on optimal selection for discrete event systems with partial observation. These methods do not address the optimal sensor placement problem which we address in this paper. Also, our method is different from the above cited works as we propose a performance metric for accomplishing the given surveillance task in order to find the optimal number of sensors.

Atrey et al [1] discuss the problem of selecting the most informative subset from the available set of media streams at any particular time instant. They select an optimal subset of streams by using dynamic programming approach for a given task to eliminate the cost of processing redundant and less informative data. The problem which we address is different from [1] as we aim to find the optimal set of sensors before building the system, whereas [1] eliminates some of the processing cost after building the system assuming that there is a lot of redundancy in the system. Also, [1] does not consider the issue of placement of sensors.

In the context of wireless sensor networks, Pahalawatta et al [8] propose to solve the problem of optimal sensor selection by maximizing the information utility gained from a set of sensors subject to a constraint on the average energy consumption in the network. Their method is not applicable to our problem as their main concern is energy consumption and the information gain is designed accordingly. A sensor placement algorithm for optimizing the coverage area have been reported in [3]. [3] assumes that the probability of detecting a target by using a sensor varies exponentially with the distance between target and the sensor. This method does not consider the notion of 'directionality'. So, it cannot be used to model the effect of PTZ cameras.

Erdem et al [4] have described an approach for placement of multiple cameras in a polygonal space using 'reasonable' assumptions for real-life cameras. Mittal et al [6] have also described a method for determining the optimal number of cameras and their placement to monitor any given premises. However, both these works do not consider direction of image captured as part of their suitability metrics. This is important as it is often necessary to obtain images in one direction (e.g. frontal direction for face/object recognition etc.) and not the other. Also, they consider only *static* cameras while we consider PTZ cameras as well as motion-sensors. On the other hand, Wren et al [9] utilize a swarm of low cost motion sensors for automatically calibrating PTZ cameras that are undertaking surveillance tasks. Their work however does not deal with finding optimal number and position etc. for the sensors.

On the whole we realize that while optimal sensor placement has generated reasonable research interest, the currently available methods fail to recognize the need for a directionality aware metric for optimally selecting and placing the sensors in a multimedia surveillance system. To the best of our knowledge, this paper is the first to address this issue.

3. PROPOSED DESIGN METHODOLGY

In this section, we describe our design methodology for obtaining the optimal set of sensors and their placement for a multimedia surveillance system. The following problem has been considered for modeling: Given l types of sensors to be employed in a convex surveyed region; the objective is to find the optimal set of sensors and their placement in order to accomplish a task with a specified performance σ. Non-convex regions can be tackled by defining a suitable coverage function (e.g. visibility for cameras) for each sensor, but in this paper, we restrict our focus only to convex regions and would like to extend it to non-convex regions in the future.

Our proposed design methodology for determining the optimal set of sensors and their placement is as follows -

Step 1 Obtain the performance metric (we denote it by η) for accomplishing the given task as a function of sensors and their placement.

Step 2 Determine the all combinations of sensors along with their placement for which the performance metric exceeds or equal to the required performance σ.

Step 3 Determine the cost of each combination and output the combination with least cost.

Let us assume that the given task can be divided into q subtasks. To obtain the performance metric (in step 1) as a function of sensors and their placement, the first step is to obtain a performance matrix of dimension $l \times q$ such that $(i,j)^{th}$ element of this matrix represents the accuracy with which i^{th} sensor can perform the j^{th} subtask. We can approximately determine this performance matrix based on our prior knowledge about various types of sensors. The next step is to decide an interaction strategy among sensors using this performance matrix. The interaction strategy should be such that the assignment of subtask(s) to particular sensor(s) results in a maximum overall performance. If we consider, for example, j^{th} column of the performance matrix, it indicates the accuracy with which j^{th} subtask can be performed by various types of sensors. Once we prioritize the subtasks, they can be assigned to the sensors depending on their priority. The final step is to model the effect of individual sensors (based on the subtask(s)) to obtain the final performance metric. Once the performance metric is obtained, we follow steps 2 and 3 to find the optimal set of sensors and their placement, as will be described in section 3.3. Please note that we do not consider the computational costs in our analysis as we assume sensor placement and selection to be offline problems.

Though the proposed design methodology is generic, we demonstrate its utility for a specific case. As mentioned earlier, we consider a surveillance task of 'capturing the frontal part of a symmetric object' in a convex surveyed region. This is because the performance metric η that we derive for

this task can be applied to other tasks such as object detection and tracking etc. This task has two sub-tasks – object localization and image capture. Two types of sensors ($l = 2$) are considered which are (PTZ)infrared cameras and active motion sensors. The localization sub-task can be performed by both types of sensors while the image capture sub-task can be performed by the camera alone. Since cameras are required in this system, we first develop the performance metric for cameras only. It is interesting to note that the performance metric which includes the effect of only cameras can be used to design an optimal camera surveillance system.

As the first sub-task task is to capture the frontal part of a symmetric object, the performance metric η could be the average probability of capturing the frontal part of a symmetric object at a particular time instant. Initially, we describe our mathematical model for obtaining the performance metric η when n number of (PTZ) infrared cameras are placed in a fixed configuration in subsection 3.1. Then we discuss optimal camera placement based on this metric in subsection 3.2. In subsection 3.3, we extend this performance metric to include the effect of the active motion sensors. Finally, we discuss about the optimal selection and placement of sensors based on this performance metric in the same subsection.

3.1 Obtaining the performance metric

We make the following assumptions while deriving the performance metric for n cameras -

1. The surveyed region is convex.
2. The symmetric object can be captured if its centroid lies within the conical FOV of cameras.
3. If half or more than half of the frontal part of the symmetric object is captured by any one of the cameras, then due to the symmetry, the frontal part of an object can be obtained.

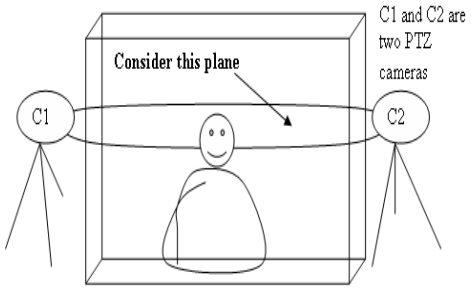

Figure 1: A typical surveillance setup

Consider a plane which is parallel to the floor and at the same vertical height as that of the cameras as shown in the figure 1. A human face is shown in the figure 1, but it could be any other symmetric object. The top view of this plane is as shown in the figure 2. Though the actual centroid of an object may not lie on the considered plane due to the variability in pose etc, most of the times the field of view (FOV) of the cameras is enough to capture the object. In this case,

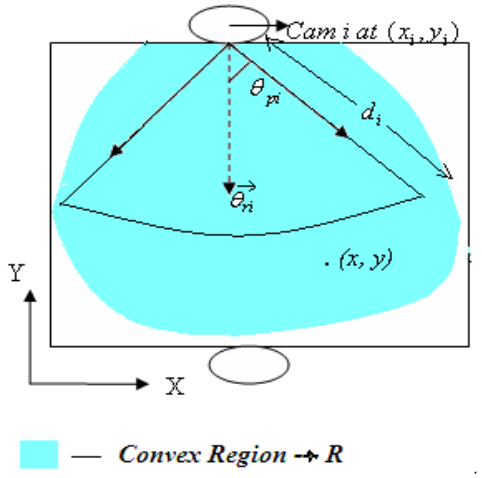

Figure 2: Top view and camera parameters

the cameras capture a slightly distorted object due to the angle of projection of an object onto the camera plane. We neglect this effect and assume that the centroid of an object lies on the considered plane for analysis. Also, in practice FOV gets affected by changes in the zoom parameter of the camera, but we neglect it as of now for the ease of modeling.

Since the surveyed region is convex, the set of all object centroid locations on this plane forms a convex region R (shaded region in figure 2). We now derive an expression for the probability of capturing a frontal part of the symmetric object if its centroid is at a location say $(x, y) \in R$ as in figure 2. This analysis will not impose any restriction on the orientation of the object. We represent the orientation of a symmetric object using a random variable which is distributed uniformly in the range $[0, 2\pi)$. This is intuitively satisfying because any orientation angle for object is equally likely. The idea is to find a set of orientation angles of an object having centroid at $(x, y) \in R$ for which the frontal part of an object can be captured by at least one of the n cameras and then determine the probability of realizing this set. By assumption 3, if we capture half or more than half of the frontal part of an object it implies that due to symmetry the total frontal part of an object can be obtained.

The parameters associated with the i^{th} camera ($1 \leq i \leq n$) are as follows:

- Location : (x_i, y_i) (on the boundary only)
- Zoom : d_i
- Reference direction :$\theta_{ri} = arg(\vec{\theta_{ri}}), 0 \leq \theta_{ri} < 2\pi$.
- Maximum pan angle : $\theta_{pi}, (> 0)$

The zoom parameter indicates the maximum distance that the i^{th} camera can focus, and the maximum pan angle indicates the maximum pan allowed in either positive or negative direction about the reference direction as shown in the figure 2. In this analysis, it is assumed that the parameter maximum pan angle includes the effect of field of view of the camera i.e. $\theta_{pi} = \theta_{pi,orig} + (FOV)/2$, where $\theta_{pi,orig}$ is the actual maximum pan angle of the camera.

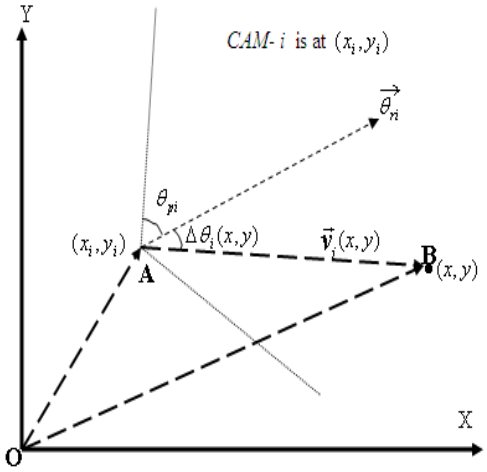

Figure 3: Directions of various vectors

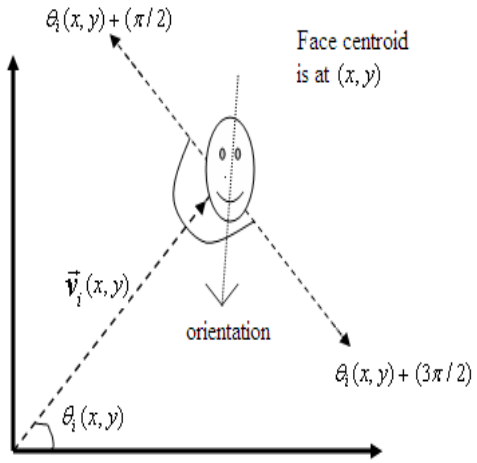

Figure 4: Face orientation

We define the *characteristic function* $I_i(x,y)$ for the i^{th} camera for all points $(x,y) \in R$ as:

$$I_i(x,y) = \begin{cases} 1, & \text{if } i^{th} \text{ camera can focus on } (x,y) \\ 0, & \text{otherwise} \end{cases}$$

and it can be written as $I_i(x,y) = I_{i1}(x,y) \times I_{i2}(x,y)$, where $I_{i1}(x,y) = U(d_i^2 - [(x-x_i)^2 + (y-y_i)^2])$ and $I_{i2}(x,y) = U(\theta_{pi} - \Delta\theta_i(x,y))$, $U(.)$ is the unit step function. $\Delta\theta_i(x,y)$ is the angle difference between the reference direction vector $(\vec{\theta_{ri}})$ of the camera and the vector $\vec{V_i}(x,y)$ as shown in the figure 3.

The characteristic function $I_i(x,y)$ essentially describes whether the object's image can be captured by camera i at point (x,y) or not. The function $I_{i1}(x,y)$ indicates the distance constraint imposed by the zoom of the camera and $I_{i2}(x,y)$ indicates the pan angle constraint. The vector from i^{th} camera to the object centroid at (x,y) is represented using $\vec{V_i}(x,y)$ and can be found using the Δ^{le} law of addition. Consider Δ^{le} `OAB` in the figure 3,
$(x_i, y_i) + \vec{V_i}(x,y) = (x,y) \Rightarrow \vec{V_i}(x,y) = (x - x_i, y - y_i)$.
Let us define $\theta_i(x,y) = arg(\vec{V_i}(x,y))$, $0 \leq \theta_i(x,y) < 2\pi$, as indicated in figure 4 for some (x,y) (here symmetric object is shown to be a face). As stated earlier, the orientation of a symmetric object is represented using a random variable θ which is distributed uniformly in the range $[0, 2\pi)$. According to the assumption 3, the i^{th} camera can capture the frontal part an object having centroid at (x,y) whenever the orientation angle of an object $\theta \in S_i(x,y)$. Figure 4 shows a specific case. $S_i(x,y)$ is expressed as -

$$S_i(x,y) = \{\theta_i : \theta_i(x,y) + \pi/2 \leq \theta_i < \theta_i(x,y) + 3\pi/2\} \bmod 2\pi$$

which represents the set of all orientation angles for an object having centroid at (x,y) for which i^{th} camera can capture the frontal part of an object. If the object is such that the frontal part of it can be obtained from any of its captured images (independent of its orientation) then the analysis becomes simple and one has to merely maximize the coverage area. This is not true for objects like human and animal faces as shown in the figure 4. Therefore, we need to do the following analysis. Let us define $P_i(x,y) = Prob\{\theta \in S_i(x,y)\}$. Hence the probability of capturing the frontal part of an object having centroid at (x,y) using the i^{th} camera is given by $P_i(x,y) \times I_i(x,y)$.
Let $P^n(x,y)$ denote the probability of capturing an object having centroid at (x,y) and with n number of cameras arranged in any fixed configuration.

3.1.1 Single camera case

Recall that $I_1(x,y)$ indicates whether camera 1 can focus on (x,y) or not. Hence, in this case, $P^1(x,y) = I_1(x,y) \times P_1(x,y)$

3.1.2 Dual camera case

We know that $P(X \bigcup Y) = P(X) + P(Y) - P(X \bigcap Y)$, where X and Y are any two events.

Case 1: When both the cameras are able to focus on (x,y)

$$P^2(x,y) = Prob\left\{\theta \in \left[S_1(x,y) \bigcup S_2(x,y)\right]\right\}$$

$$= Prob\{\theta \in S_1(x,y)\} + Prob\{\theta \in S_2(x,y)\}$$

$$- Prob\left\{\theta \in \left[S_1(x,y) \bigcap S_2(x,y)\right]\right\}$$

$$= P_1(x,y) + P_2(x,y) - P_{12}(x,y) \text{ , where}$$

$$P_{12}(x,y) = Prob\left\{\theta \in \left[S_1(x,y) \bigcap S_2(x,y)\right]\right\}$$

and denotes the probability of capturing the frontal part of an object having centroid at (x,y) by both the cameras.
Case 2: When only one of the cameras is able to focus on (x,y)

$$P^2(x,y) = Prob\{\theta \in S_i(x,y)\} = P_i(x,y)$$

where only camera i can focus on (x,y), $i = 1$ or 2
Case 3: Either of the cameras can't focus on (x,y)
$P^2(x,y) = 0$

The above all cases can be compactly written as

$$P^2(x,y) = I_1(x,y) \times P_1(x,y) + I_2(x,y) \times P_2(x,y)$$

$$-I_1(x,y) \times I_2(x,y) \times P_{12}(x,y)$$

Since the random variable θ is uniformly distributed in the range [0, 2π), the above expression reduces to

$$P^2(x,y) = (1/2) \times [I_1(x,y) + I_2(x,y)]$$

$$-I_1(x,y) \times I_2(x,y) \times P_{12}(x,y) \ldots\ldots (1)$$

The point (x,y) can be anywhere on the plane and belongs to the convex set R and the characteristic function of a paricular camera describes whether that camera can focus on this point or not. Average probability of capturing the frontal part of an object at any particular time instant can be found if we know the probability density function $f(x,y)$ for an object centroid position over the convex region R. Let the average probability be η and represents the performance metric as discussed earlier.

$$\eta = \iint_R P^2(x,y)\ f(x,y)\ dx\ dy$$

Let the area of convex region R be A_R and further assume that the position (x,y) is a random variable with a uniform density (in this case, $f(x,y) = \frac{1}{A_R}$). Uniform density for the position means object can be found with an equal probability in any region of fixed total area.

$$\eta = \frac{1}{A_R} \iint_R P^2(x,y)\ dx\ dy$$

Substituting for $P^2(x,y)$ from (1),

$$= \frac{1}{A_R} \iint_R (1/2) \times [I_1(x,y) + I_2(x,y)]\ dx\ dy$$

$$-\frac{1}{A_R} \iint_R I_1(x,y) \times I_2(x,y) \times P_{12}(x,y)\ dx\ dy$$

$$\eta = \frac{0.5}{A_R} \{Volume\ under I_1(x,y) + Volume\ under I_2(x,y)\}$$

$$-\frac{1}{A_R} \iint_A P_{12}(x,y)\ dx\ dy \ldots\ldots (2)$$

where, A : Area where both the cameras can focus.
(i.e., Set of all (x,y) under Case 1)

3.1.3 More than two cameras

In this section we extend the performance metric to the n camera case. As mentioned earlier, $P^n(x,y)$ denotes the probability of capturing the frontal part of an object having centroid at (x,y) and with n number of PTZ cameras in a fixed layout. If (x,y) is such that all cameras are able to focus on this point then expression for $P^n(x,y)$ is given by:

$$P^n(x,y) = Prob\left\{\theta \in \left[S_1(x,y) \bigcup S_2(x,y) \ldots \bigcup S_n(x,y)\right]\right\}$$

Since we know how to deal with two cameras, initially we start with two cameras. After determining the effect of first two cameras, we add one more camera to find its effect. Note that the order in which we add cameras to the existing configuration has no effect on the final performance metric as the union operator is associative. This process of adding a new camera to the existing system is repeated till we include all the cameras. The algorithmic approach is described below.

Algorithm 1: To determine $P^n(x,y)$
Inputs:
Sets: $S_i(x,y)$, $i = 1, 2, \ldots n$
Probabilities: $P_i(x,y)$, $i = 1, 2, \ldots n$
Charcteristic functions: $I_i(x,y)$, $i = 1, 2, \ldots n$
Initialize:
$A \leftarrow S_1(x,y)$ and $B \leftarrow S_2(x,y)$
$p_1 \leftarrow P_1(x,y)$ and $p_2 \leftarrow P_2(x,y)$
$i_1 \leftarrow I_1(x,y)$ and $i_2 \leftarrow I_2(x,y)$
for $j = 3$ to n
Compute:
$p = i_1 \times p_1 + i_2 \times p_2 - i_1 \times i_2 \times p_{12}$
where $p_{12} = Prob\{\theta \in A \bigcup B\}$
Update sets:
if $i_1 = 1$ and $i_2 = 1$
then $A \leftarrow A \bigcup B$
if $i_1 = 0$ and $i_2 = 1$
then $A \leftarrow B$
if $i_1 = 0$ and $i_2 = 0$
then $A \leftarrow \phi$
Update probabilities:
$p_1 \leftarrow p$ and $p_2 \leftarrow P_j(x,y)$
Update characteristic functions:
$i_1 \leftarrow \max(i_1, i_2)$ and $i_2 \leftarrow I_j(x,y)$
end for
$P^n(x,y) = i_1 \times p_1 + i_2 \times p_2 - i_1 \times i_2 \times p_{12}$
Output $P^n(x,y)$

Once we know $\{P^n(x,y),\ \forall (x,y) \in R\}$, the average probability η can be found by integrating and averaging over the entire convex region as discussed in subsection 3.1.2. This average probability represents the performance metric for the n camera case. Optimal camera placement is obtained by maximizing η with respect to the camera placement.

3.2 Optimal camera placement

The performance metric η derived in section 3.1 is used to determine the optimal camera placement. The optimal camera placement refers to the placement of cameras which gives the maximum performance metric. We determine the η by displacing the cameras along the perimeter and then find the placement which gives the maximum η. The reference direction for any camera is chosen such that maximum volume is included under the corresponding characteristic function. Simulation results for the two camera case are presented in the results section.

Algorithm 2: (optimal camera placement, n cameras)
Inputs: Number of cameras, specifications
Step 1: Choose the optimal reference direction when a camera is placed at a particular point on the perimeter.
Step 2: Determine η for this particular camera placement.
Step 3: Repeat the above steps by displacing the individual cameras along the perimeter.
Step 4: Pick the placement of cameras which gives the maximum η and output it.

3.3 Modeling of the effect of motion sensors

We derived a performance metric as a function of cameras and their placement in section 3.1. In this section, we ex-

tend this performance metric to include the effect of motion sensors (described in section 4.1).

The motion sensor grid (section 4.1) in our proposed framework is used for localizing an object (section 4.2). Object could be anywhere not cutting the adjacent motion sensors beam after cutting the intersection/grid point. In this case, the field of view of the camera (refer to figure 5) plays a vital role in determining the average probability of capturing an object. This is because when camera is focusing on the grid point, if an object centroid lies outside the FOV of the camera then it is not possible to capture that object. Object

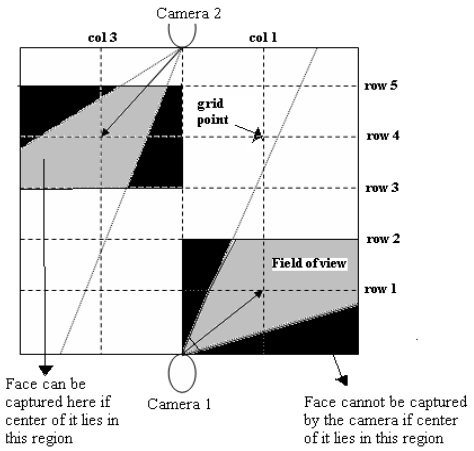

Figure 5: Importance of FOV of the camera

can be captured only if its centroid lies within the FOV of the camera. In other words, the FOV of the camera allows system to have some uncertainty in localizing an object and further facilitates the capture of the object. The allowed uncertainty increases as with the FOV. The region of uncertainty (rectangle) associated with a grid point (refer section 4.2) can be reduced by increasing the number of motion sensors in a system. Hence there is a tradeoff between the FOV of the camera and the number of motion sensors used for localization. If the number of motion sensors is such that the uncertainty in localizing an object using the sensor grid is just equal to the allowed uncertainty due to the field of view of the camera then that number indicates the optimal number of motion sensors. Increasing the number of motion sensors beyond the optimal number does not improve the performance metric of the system.

For any particular grid point, the performance of the system is specified in terms of probability of capturing the frontal part of an object when all cameras focus on this point. We assume uniform probability density function for an object position in the region of uncertainty associated with any particular grid point and determine the performance of the system corresponding to this grid point. By assuming equal probability for the cameras to focus on any particular grid point, the final performance metric η can be obtained by taking the average of all performances corresponding to grid points. The following is the algorithm for determining the performance metric η for our heterogeneous sensor system.

Algorithm 3
Input: Sensors (both cameras and motion sensors) and their placement
Step 1: Consider a grid point and obtain the corresponding region of uncertainty for the object centroid position.
Step 2: Determine the average probability of capturing the frontal part of an object for this grid point.
Step 3: Repeat the above steps for all grid points and then output the average of all the probabilities (of step2).
Output: Performance metric η.

The performance metric η is a function of number of sensors and their placement. Tradeoff between the FOV of the camera and number of motion sensors allows us to select an optimal number of motion sensors for a given FOV. By fixing the number of cameras and varying the FOV, we can select an optimal combination of FOV and the associated optimal number of motion sensors. This is because as the FOV of the camera increases, the cost of the camera increases and the associated optimal number of motion sensors decreases. Hence, there exists an optimal combination of FOV for the cameras and the number of motion sensors for which the over all cost is minimum. Finally, the overall optimized heterogeneous sensor system can be obtained by minimizing the overall system cost with respect to number of cameras and the corresponding optimal combination.

3.4 Fault tolerance

Once the multimedia surveillance system is built, it is important to know how the system performance deteriorates when few components (sensors) of the system fail. We can determine the performance of the system in this case using the algorithm 3 by removing the faulty components from the inputs list. Using this algorithm, system designer can estimate the performance of the system when few components of it fail before building the actual system. Thus, this methodology also serves as a powerful failure analysis tool and can help design a system with a graceful performance degradation under sensor failure.

4. SYSTEM IMPLEMENTATION

After selecting the optimal subset of the sensors and their placement as described in section 3.3, the next step is to build a surveillance system using these sensors to accomplish the given task. We discuss the implementation details of a multimedia surveillance system in subsection 4.1. Subsection 4.2 discusses the uncertainty in localizing an object when motion sensor grid is employed for localization. Finally in subsection 4.3, we describe the use of coopetitive interaction strategy in our system.

4.1 System implementation details

For a surveillance task of capturing the frontal part of a symmetric object in a rectangular region of $6m \times 2.5m$, we considered two types of sensors - PTZ infrared cameras (Canon VC C50i) and motion sensors. We followed the design methodology and determined the number of infrared cameras and motion sensors as two and eight, respectively; and also their placement is determined as shown in figure 6. The system consists of two PTZ cameras placed at diagonally opposite corners and the eight motion sensors arranged in the form of a (2-D) 5×3 grid (figure 6).

Motion sensor consists of a transmitter and a receiver pair. The transmitter emits the IR light (source) and the corre-

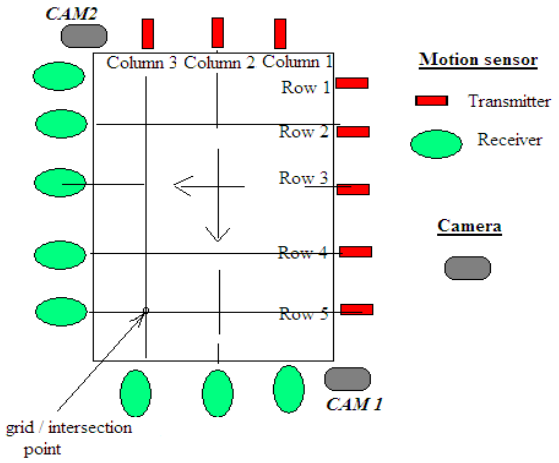

Figure 6: Motion-sensor grid and camera positions

sponding receiver detects it when placed opposite to the transmitter. Receiver cannot detect any IR light if any obstacle (or intruder) obstructs the beam (IR light). This fact is exploited when the motion sensor grid is deployed for localizing an object in a dark region. If we consider any particular motion sensor (along row or column), there are two states associated with it namely 'beam is continuous' and 'beam is discontinuous' and hence single bit is sufficient to represent these two states. Hence for eight motion sensors we require eight bits or one byte. The computer can access the current state of the motion sensor grid via its serial port operating at 115200 baud. Let us call the motion sensors along the row as 'row motion sensors' and along the column as 'column motion sensors' as shown in the figure 6. The grid points correspond to the locations where beams from the row motion sensors and column motion sensors meet. Therefore, any individual grid point can be specified by its row and column motion sensor number i.e., (r, c), where $1 \leq r \leq 5$ and $1 \leq c \leq 3$ for our system.

The cameras used in our system can be controlled by setting different pan, tilt and zoom parameters. They are operated in infrared mode to capture images of an object when there is no illumination. As shown in the figure 6, the two cameras are placed at the diagonally opposite corners and are in fact at the same vertical height.

4.2 Localization by motion sensor grid

The active motion sensor used in our system provides information such as 'something is obstructing the beam' or 'nothing is obstructing the beam'. The uncertainty in localizing an object when it obstructs the beam from any single motion sensor is that it could be obstructing the beam anywhere on the line joining the transmitter and the corresponding receiver. When the beam is continuous the uncertainty is that an object could be anywhere but not on the line joining the transmitter and the corresponding receiver. In localizing an object using the sensor grid, we assume that the object cannot cross the beam of any motion sensor in less than Δt seconds, where Δt is the polling time for the sensor grid. In other words, it is the time difference between the two consecutive sensor grid data reads. Using the past localization information and the current sensor grid status, the new localization information can be obtained i.e., $R(t) = f1(R(t-1), S(t))$ and $C(t) = f2(C(t-1), S(t))$, where, $R(t)$ and $C(t)$ represent the row and column motion senor numbers to focus at time instant t respectively. S(t) denotes the sensor data at time instant t. $f1()$ and $f2()$ show the functional dependence of $R(t)$ and $C(t)$ respectively.

It is always required to force the cameras to focus on the grid/intersection point to reduce the uncertainty in capturing an object. The following explains the different cases.

Case 1: *At time instant t, the row motion sensor with number r and the column motion sensor with number c are discontinuous.*
In this case there is no uncertainty in localizing an object and it is exactly there on the grid/intersection point (r, c).

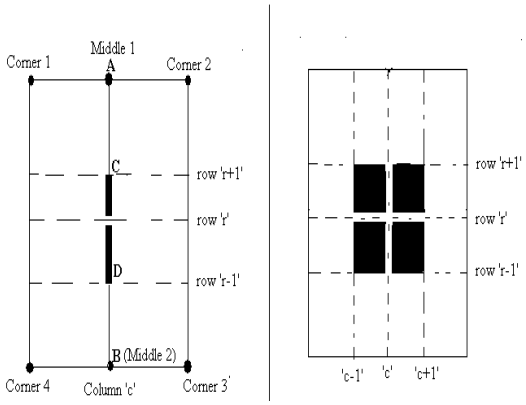

Figure 7: Uncertainty in localization

Case 2: *At time instant t, only one of the motion sensors is discontinuous (row or column).*
Let the current discontinuous motion sensor be a column motion sensor with number c (refer figure 7, left). Let the previous latest discontinuous row motion sensor number be r. As per the assumption in section 4.2, object cannot cross either row $r-1$ or $r+1$ and yet obstructing column motion sensor c. Because of the assumption, the uncertainty in localizing an object in this case is reduced from the line segment AB to the line segment CD (thick) excluding the intersection/grid point as shown in the figure 7. So it is necessary to focus the cameras on the grid/intersection point (r, c) to reduce the uncertainty of capturing an object.

Case 3: *At time t, no motion sensor is discontinuous.*
Let the latest previous discontinuous row and column motion sensors be r and c respectively. In this case an object cannot cross the row motion sensors $r-1$ and $r+1$ and similarly the column motion sensors $c-1$ and $c+1$. Hence the uncertainty region in this case is the dark region as shown in the figure 7(right)(note that the row motion sensor r and column motion sensor c are continuous). So, by focusing the cameras on the grid point (r, c), we can reduce the uncertainty of capturing an object.

4.3 Interaction strategy

In this section, we describe the interaction strategy used by our heterogeneous sensor system. The design of the in-

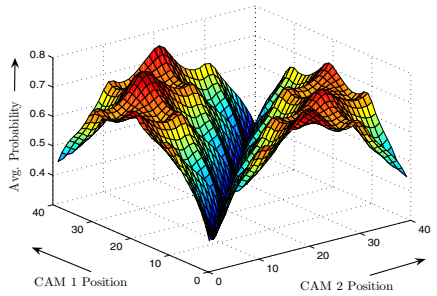

Figure 8: Performance vs. camera placement

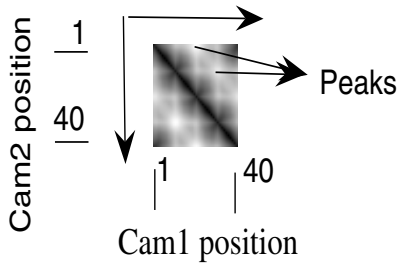

Figure 9: Image of the probability distribution

teraction strategy for multimedia system must fuse the prior knowledge about the various capabilities of sensors into the controlling algorithm to maximize the throughput. The considered task of capturing a frontal part of an object can be divided into two subtasks namely localizing an object and capturing the images at the specified localized location. Motion sensor grid can perform only the first subtask whereas infrared cameras can perform both the subtasks. Motion sensor grid performs better than cameras for the first subtask. Based on the above prior information, maximum throughput is achieved by assigning the first subtask to motion sensor grid and second subtask to infrared cameras as it is the only available sensor which can undertake this subtask. The algorithmic steps are as follows.

Algorithm 4
Step 1: Obtain the motion sensor grid data and determine the intersection point to focus.
Step 2: Steer the cameras to focus on this point.
Step 3: Capture images using both the cameras and search for the frontal part of an object.
Step 4: Repeat the above steps 1, 2 and 3 till the system captures some specified number of frontal object images.

5. RESULTS

We present in this section simulation results describing the optimal selection of heterogeneous sensors and their placement. We also show the experimental results for tracking and capturing the face of an intruder.

5.1 Optimal camera placement

5.1.1 Square region

Even though the proposed algorithm is generalized for n cameras and any arbitrary convex area, some insights can be obtained by considering a dual camera placement problem for a square area of $20m \times 20m$. Total 40 equally spaced points are considered along the perimeter of a square. The corner points on the perimeter of a square are numbered as 1, 11, 21, and 31 respectively. Maximum pan angle (θ_{pi}) is chosen to be 45 degrees and 20 for the zoom.

Figure 8 shows the performance metric(average probability of capturing the frontal part of an object) of the system as a function of cameras position along the perimeter. We can easily see the two way symmetry of this function when it is represented as an image (the intensity of any pixel is proportional to the average probability) as shown in figure 9. This is because camera positions can be swapped without changing the performance of the system and the other reason is that the considered region is a square. When both the cameras are placed at the same point then the performance of the system is same as that of the single camera placed at this point. The dark line along the diagonal in the figure 9 represents this effect. Maximum performance occurs for total 2 combinations, locations 1 and 21, 11 and 31. Since cameras can be swapped for each combination, there are total 4 peaks in the function.

The pan angle and the zoom are chosen such that maximum volume under the characteristic function is obtained for most of the camera positions along the perimeter when cameras choose the optimal reference directions. According to the equation (2) of section 3.1.2, average probability can be maximized by maximizing the volume under the characteristic functions and simultaneously minimizing the intersection region of the characteristic functions. In this example, the performance can be maximized by minimizing the intersection region of the characteristic functions. This is because parameters are chosen such that most of the combinations have the same volume (maximum) under the characteristic functions. Intersection region can be minimized by placing the cameras far apart. The two farthest points on the perimeter of a square are the end points of a diagonal. Hence equation (2) of section 3.1.2 says that place the cameras on the end points of a diagonal to maximize the performance. We got the same results through our simulations as discussed above.

5.1.2 Irregular pentagon

To study the generalizability, we investigated a more complex geometry and thus considered the case of a convex surveyed area as shown in the figure 10 and analyzed the dual camera placement problem.

Maximum pan angle (θ_{pi}) and zoom are chosen to be 55 degree and 24, respectively for both the cameras. The perimeter of the area is divided into 60 equal parts (one part is $2m$). By displacing the cameras along the perimeter we obtained the performance of the system as shown in the figure 11. The combination 48 and 11 gave the maximum performance of 0.6. Figure 10 shows the performance metric as a function of spacial location (i.e., $P^2(x,y), \forall (x,y) \in R$) when cameras are placed optimally. Note that intensity of the any pixel (x,y) is proportional to the probability $P^2(x,y)$.

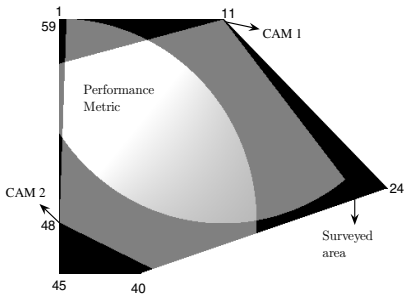

Figure 10: Surveyed region

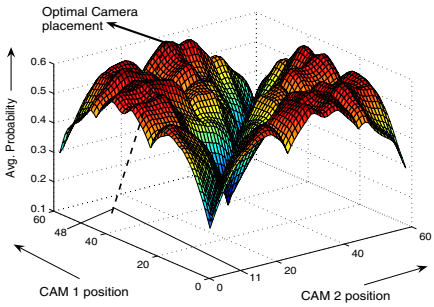

Figure 11: Performance vs. camera placement

5.2 Optimal selection of sensors

We consider the same square area of $20m \times 20m$ and the two cameras are placed optimally as per the above section 5.1.1. Figure 12 shows the trade off plots for different maximum pan angles. Number along the $Y - axis$ represents the $(number\ of\ motion\ sensors - 4)/2$ and the FOV along the $X - axis$ is $(field\ of\ view\ in\ degrees)/10$. We can see from the figure 12 that there is no increase in the average probability beyond a particular value of approximately 0.63 (top surface) and in fact it is saturating. The optimal combination for achieving the performance of 0.63 is 12 motion sensors(6×6) and FOV of 40 degrees for both the cameras when the total pan angle of an individual camera is 60 degrees ($\theta_{pi} = 30$ degrees). This is optimal because

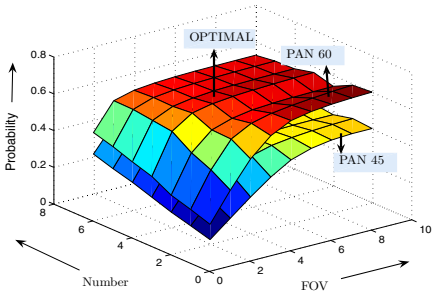

Figure 12: Tradeoff plots

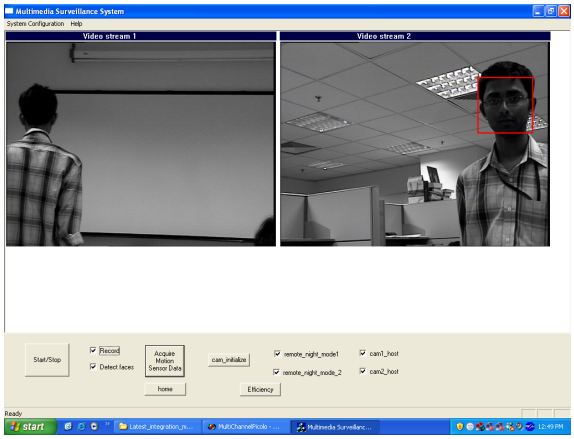

Figure 13: Test bed for our experiments

Table 1: Effect of camera placement

Camera placement	Face capturing ratio (%)
Cam1-Middle1, Cam2-Middle2	42
Cam1-Middle1, Cam2-Corner3	51
Cam1-Corner1, Cam2-Corner3	69
Cam1-Corner1, Cam2-Corner4	38
Cam1-Corner1, Cam2-Corner1	27
Cam1-Middle1, Cam2-Middle1	17

this combination has got both minimum number of motion sensors and minimum FOV out of all feasible combinations (which give performance greater than or equal to 0.63).

5.3 Tracking results

In this section, we present face tracking results of the system described in section 4.1. To track and further capture the frontal face of an intruder, cameras parameters like pan, tilt and zoom need to be adjusted based on the localization information obtained from motion sensors. Such an interaction strategy between sensors allows the system to react and track an intruder efficiently. For example consider figure 14 where few images captured by both the cameras of a surveillance system for a particular camera placement are shown. Since localization is done by the motion sensor grid, cameras are able to react and track an intruder even if no face is being detected in the captured frames. This can be observed from images (g), (g'), (h) and (h') of figure 14. Surveillance systems consisting of only cameras cannot track in this case.

Table 1 summarizes the effect of camera placements on the 'successful face capturing' ratio. We define 'successful face capturing' ratio of the number of frames captured with frontal facial data to the total number of frames captured for each camera. In our experiments, we considered a fixed motion trajectory that passes through all the grid points and obtained 100 frame images per camera for each camera placement. Total 6 points were chosen (i.e.,Corner1-4 and Middle1-2) along the perimeter for the camera position as shown in the left image of figure 7. The experimental results show that maximum accuracy of 69 percent is obtained when cameras are placed in diagonally opposite corners. Note that, equation (2) in section 3.1 also suggests the same placement for obtaining the maximum performance.

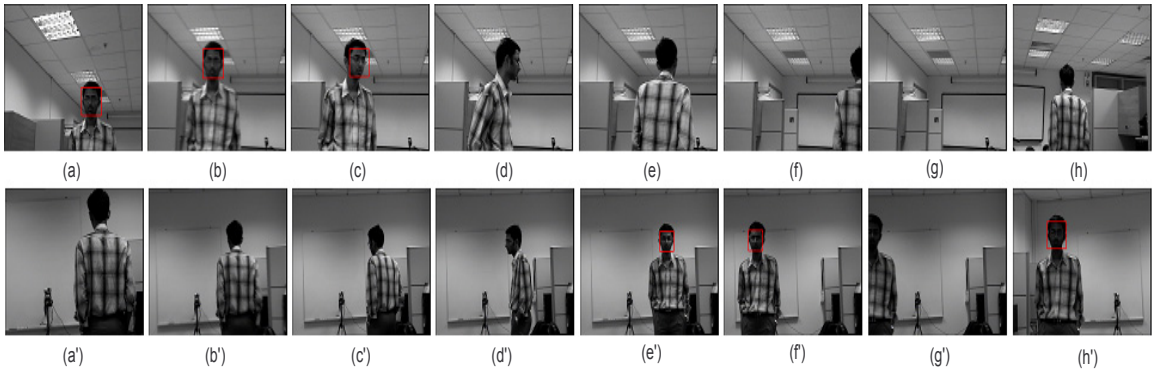

Figure 14: Tracking results: (a)-(h) Camera 1 images, (a')-(h') Camera 2 images

Table 2: Performance metric as a function of sensors

Sensor grid	performance metric (2 cameras)	performance metric (1 camera)
3×2	0.4278	0.2210
4×2	0.5049	0.2634
4×3	0.5266	0.2784
5×3	0.5495	0.2913

5.4 Fault tolerance

For a rectangular region of $20m \times 10m$, Table 2 lists the performance metric discussed in section 3.3 for various combinations of sensors. Using this table, we can estimate the degradation in system's performance when few sensors fail without building the actual system. In table 2, $m_1 \times m_2$ under the sensor grid represents the number of sensors along the length $(20m)$ and width$(10m)$ respectively. The individual camera parameters are chosen as follows:
pan angle $= \pi/3$, zoom $= 14$, FOV$=\pi/6$. Cameras are assumed to be placed along the length as shown in the figure6. Thus, the system designer can estimate the performance even before the system is built. If the estimated performance under failure is unacceptable, the designer can choose to add more sensors (albeit at a higher cost) to ensure a minimum performance level for sensitive applications.

6. CONCLUSIONS

In this paper, we have proposed a new performance metric for accomplishing the given surveillance task using heterogenous sensors. We have presented a novel design methodology based on this metric that can help obtain the optimal combination of sensors and further their placement in a given surveyed area. Simulation results have shown the power of these algorithms in obtaining the optimal combination. Future work includes the modeling the effect of temporal component for a dynamically changing task and inclusion of other sensors like microphones etc into the mathematical analysis. We also intend to conduct rigorous experimentation with more than two cameras and handle multiple intruders as well as occlusion effects. Finally, we would like to extend this work to non-convex regions.

7. REFERENCES

[1] P. K. Atrey and M. Kankanhalli. Goal based optimal selection of media streams. In *IEEE International Conference on Multimedia and Expo*, pages 305–308, Amsterdam, The Netherlands, July 2005.

[2] R. Debouk, S. Lafortune, and D. Teneketzis. On an optimal problem in sensor selection. In *Discrete Event Dynamic Systems: Theory and Applications*, volume 12, pages 417–445, March 2002.

[3] S. S. Dhillon and K. Chakrabarty. Sensor placement for effective coverage and surveillance in distributed sensor networks. In *IEEE Wireless Communications and Networking Conference*, pages 1609–1614, New Orleans, USA, March 2003.

[4] U. M. Erdem and S. Sclaroff. Optimal placement of cameras in floorplans to satisfy task requirements and cost constraints. In *International Workshop on Omnidirectional Vision*, Prague, Czech Republic, May 2004.

[5] S. Jiang, R. Kumar, and H. E. Garcia. Optimal sensor selection for discrete event systems with partial observation. In *IEEE Transactions on Automatic Control, vol. 48*, pages 369–381, March 2003.

[6] A. Mittal and L. Davis. Visibility analysis and sensor planning in dynamic environments. In *European Conference on Computer Vision*, Prague, Czech Republic, May 2004.

[7] Y. Oshman. Optimal sensor selection strategy for discrete-time state estimators. In *IEEE Transactions on Aerospace and Electronic Systems, vol. 30*, pages 307–314, April 1994.

[8] P. Pahalawatta, T. N. Pappas, , and A. K. Katsaggelos. Optimal sensor selection for video-based target tracking in a wireless sensor network. In *IEEE International Conf. on Image Processing*, Singapore, October 2004.

[9] C. R. Wren, U. Erdem, and A. Azarbayejani. Automatic pan-tilt-zoom calibration in the presence of hybrid sensor networks. In *ACM International Workshop on Video Surveillance and Sensor Networks*, Singapore, Nov 2005.