Sensor Planning for Automated and Persistent Object Tracking with Multiple Cameras

Yi Yao, Chung-Hao Chen, Besma Abidi, David Page, Andreas Koschan, Mongi Abidi
Imaging, Robotics, and Intelligent Systems (IRIS) Lab, University of Tennessee, Knoxville

Abstract

Most existing camera placement algorithms focus on coverage and/or visibility analysis, which ensures that the object of interest is visible in the camera's field of view (FOV). However, visibility, a fundamental requirement of object tracking, is insufficient for persistent and automated tracking. In such applications, a continuous and consistently labeled trajectory of the same object should be maintained across different cameras' views. Therefore, a sufficient overlap between the cameras' FOVs should be secured so that camera handoff can be executed successfully and automatically before the object of interest becomes untraceable or unidentifiable. The proposed sensor planning method improves existing algorithms by adding handoff rate analysis, which preserves necessary overlapped FOVs for an optimal handoff success rate. In addition, special considerations such as resolution and frontal view requirements are addressed using two approaches: direct constraint and adaptive weight. The resulting camera placement is compared with a reference algorithm by Erdem and Sclaroff. Significantly improved handoff success rate and frontal view percentage are illustrated via experiments using typical office floor plans.

1. Introduction

With the increased scale and complexity involved in most practical surveillance applications, it is almost impossible for any single camera (either fisheye or PTZ) to fulfill persistent tracking and monitoring tasks with an acceptable degree of continuity and/or reasonable accuracy. Systems with multiple cameras find extensive use in surveillance applications. The need for sensor planning comes naturally when the question of how to place multiple cameras to fulfill given tasks with given performance requirements arises.

In literature, most sensor planning algorithms are proposed for such applications as 3D object inspection and reconstruction. Roy *et al.* reviewed existing sensor planning algorithms for 3D object reconstruction [1] and proposed an online scheme using a probabilistic reasoning framework for next-view planning and object recognition [2]. Wong *et al.* defined a metric evaluating the unknown information in each group of potential viewpoints and used it in the search of the next best view for 3D modeling [3]. Yous *et al.* designed an active scheme for multiple PTZ camera assignment so that each camera observes a specific part of a moving object, mainly pedestrians, and achieves the best visibility of the whole object [4]. Sensor planning for surveillance systems also received increasing attention in recent years [5-7]. Cameras are placed such that a full or specified coverage of the environment or object is achieved. A probabilistic camera planning framework with visibility analysis was proposed by Mittal and Davis [8]. Erdem and Sclaroff defined different types of coverage problems and developed corresponding solutions using perspective cameras [9].

The conventional requirements in sensor planning, such as coverage and visibility, are unable to ensure a persistent and automated tracking in real-time surveillance systems. Sufficient amounts of overlap between the FOVs of adjacent cameras should be reserved so that consistent labeling and camera handoff can be executed successfully. However, coverage and overlapped FOVs are at two opposite ends. Given the same camera configuration, an increase in overlapped FOVs leads to a decrease in coverage. Therefore, an optimal balance between coverage and overlapped FOVs is to be found via sensor planning. The optimal balance requires a maximum increase in handoff success rate at the cost of a minimum decrease in coverage using the same number of cameras.

We refer to the necessary overlapped FOVs as handoff safety margin and design an observation measure (OM) to differentiate it from other visible areas in the camera's FOV. Based on this safety margin, we develop sensor planning algorithms balancing the tradeoff between overall coverage and adequate overlap. Furthermore, variations, including direct constraint and adaptive weight approaches, are introduced for special considerations of resolution and frontal view requirements. We then compare the efficiency of our algorithms with the scheme proposed by Erdem and Sclaroff [9] under three criteria: coverage, handoff success rate, and frontal view percentage using typical office floor plans, an example of which is shown in Figure 1.

The major contributions of this paper are listed as follows. (1) An observation measure is designed for perspective cameras to describe the suitability of tracking and to define the handoff safety margin. (2) A general sensor planning algorithm for persistent and automated tracking is developed to secure sufficient handoff margins. (3) Special
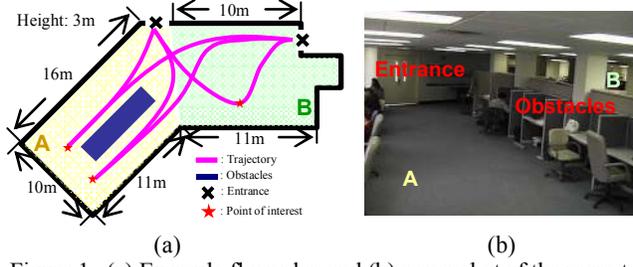
Figure 1. (a) Example floor plan and (b) a snapshot of the room to surveil.

considerations such as resolution and frontal view are addressed by two types of solutions: direct constraint and adaptive weight.

The remainder of this paper is organized as follows. The problem definition is given in Section 2. Section 3 defines the observation measure. Our sensor planning algorithms are described in Section 4. Section 5 demonstrates our experimental results and comparisons with the reference algorithm. Section 6 concludes this paper.

2. Problem definition

When formulated mathematically as an optimization process, sensor planning algorithms fall into two categories: (1) the search for the maximum coverage given a fixed total cost or number of cameras and (2) the search for the minimum cost or number of cameras for a full or designated coverage. In this paper, we refer to (1) and (2) as Type 1 and Type 2 problems, respectively.

Assuming that a polygonal floor plan is represented as an occupancy grid, a binary vector $\mathbf{b}$ can be obtained by letting $b_i = 1$ if the corresponding grid can be seen by at least one camera and $b_i = 0$ otherwise. We construct a binary matrix A with $a_{ij} = 1$ if the ith grid is covered by the jth camera configuration. Each camera configuration specifies one combination of camera intrinsic and extrinsic parameters, including camera's focal length f, pan/tilt angle θ_P / θ_T, and position T_C. The following relation holds: $b_i = 1$ if $b_i' > 0$; and $b_i = 0$ otherwise, with $\mathbf{b'} = A\mathbf{x}$ where the solution vector $\mathbf{x}$ is a set of chosen camera configurations with the corresponding elements x_j being one.

Let the cost associated with the jth camera configuration be ω_j. Given the maximum cost C_{max}, the Type 1 sensor planning problem can be described by:

$$\max \sum_i b_i, \text{ subject to} \sum_j \omega_j x_j \le C_{max} . \qquad (1)$$

Given a specified coverage vector $\mathbf{b}_o$ or a minimum coverage percentage C_{min}, the Type 2 problem can be modeled as:

$$\min \sum_j \omega_j x_j, \text{ subject to } A\mathbf{x} \ge \mathbf{b}_o , \qquad (2)$$

or

$$\min \sum_j \omega_j x_j, \text{ subject to } \sum_i b_i \ge C_{min} . \qquad (3)$$

The Type 2 problem with specified coverage was addressed and solved using binary programming in [9].

3. Observation measure

To describe the observation of a tracked target in addition to visibility, we consider the resolution M_R and the distance to the edges of camera's FOV M_D. From a viewer's point, visibility is a fundamental requirement. Herewith, the viewer includes not only operators but also successive processing such as consistent labeling and face/object recognition. An observation with different detail levels affects the performance of these algorithms. For example, a frontal face image with an inter-ocular distance of 60 pixels is recommended by a well-known face recognition engine FaceIt® for a face to be automatically recognized [10]. The M_R component is designed to evaluate a valid observation for the viewer. For a persistent object tracking and smooth camera handoff, the tracked target should be at a reasonable distance from the edges of the camera's FOV. The M_D component considers the safety margin before the object falls out of the camera's FOV.

3.1. Definition

To begin our study, the camera and world coordinates are defined and illustrated in Figure 2. A point $P = [X \ \ Y \ \ Z]^T$ in world coordinates is projected onto a point $\mathbf{p} = [x' \ \ y' \ \ z']^T$ in camera coordinates by:

$$\begin{bmatrix} x' \\ y' \\ z' \end{bmatrix} = \begin{bmatrix} \cos\theta_T & 0 & -\sin\theta_T \\ 0 & 1 & 0 \\ \sin\theta_T & 0 & \cos\theta_T \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos\theta_P & \sin\theta_P \\ 0 & -\sin\theta_P & \cos\theta_P \end{bmatrix} \begin{bmatrix} Z - T_Z \\ X - T_X \\ Y - T_Y \end{bmatrix}, \qquad (4)$$

with $T_C = [T_X \ \ T_Y \ \ T_Z]^T$. Assuming zero skew, unit aspect ratio, and zero image center, the projected point in the image plane is given by: $\begin{cases} x = fx'/z' \\ y = fy'/z' \end{cases}$.

Letting $Z=0$ (in the ground plane), the target depth $\hat{z}'$ can be estimated by:

$$\hat{z}' = \frac{-T_Z}{x/f \cos\theta_T + \sin\theta_T} , \qquad (5)$$

and the resolution component is then expressed as $M_R = \alpha f / \hat{z}'$, where α is a normalization coefficient. However, when the target is at a close distance, this relation is not entirely valid, especially when part of the target falls out of the camera's FOV. Therefore, the above definition is adjusted by:

$$M_R = \begin{cases} \alpha f / \hat{z}' & \hat{z}' > -T_Z / \tan\theta_T \\ \dfrac{\alpha f}{(\hat{z}' + T_Z / \tan\theta_T)^2 - T_Z / \tan\theta_T} & \hat{z}' \le -T_Z / \tan\theta_T \end{cases} . \qquad (6)$$
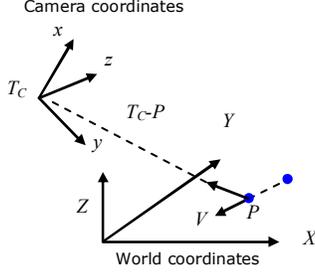
Figure 2. Illustration of the camera and world coordinates.

In practice, for a better observation and to reserve enough computation time for handoff, the target should remain at a distance from the edges of the camera's FOV. Moreover, this margin distance is affected by the target depth. When the target is at a closer distance, its projected image undergoes larger displacement in the image plane. Therefore, a larger margin should be reserved. In our definition, different polynomial powers are used to achieve varying decreasing rates of the M_D component as the object of interest approaches the edges of the camera's FOV. The M_D is then given by:

$$M_D = \beta \left[\left(|x| - N_{row}/2 \right)^2 + \left(|y| - N_{col}/2 \right)^2 \right]^{\beta_1 \hat{z}' + \beta_0}, \quad (7)$$

where N_{col} and N_{row} denote the image's width and height, β is a normalization weight, and coefficients β_1 and β_0 are used to adjust the polynomial power.

The observation measure for a perspective camera is then given by:

$$S = \begin{cases} w_R M_R + w_D M_D & (x,y) \in \Pi \\ -\infty & otherwise \end{cases}, \quad (8)$$

where w_R and w_D are importance weights and Π denotes the image plane. Note that no explicit visibility component is present in (8). Instead, the visibility analysis is included in the definition of M_R and M_D.

3.2. Handoff safety margin

A failure threshold S_F and a trigger threshold S_T are derived to define three disjoint regions: (1) invisible area with $S_{ij} < S_F$ where S_{ij} represents the OM value of the ith grid observed by the jth camera configuration, (2) visible area with $S_{ij} \geq S_T$, and (3) handoff safety margin with $S_F \leq S_{ij} < S_T$. The failure threshold S_F segments the invisible areas and is used for coverage analysis. The trigger threshold S_T separates the visible areas and handoff margins. It is introduced for handoff rate analysis, where the overlapped FOVs between adjacent cameras are optimized. The trigger threshold S_T is given by $S_T = S_F + \mu V_m T_H$ where V_m represents the average moving speed of the object of interest, T_H denotes the average

duration for a successful handoff, and μ is a conversion scalar.

The individual and combined effects of M_R and M_D become evident when we study the contours of OM projected onto the ground plane. In Figure 3, the black solid lines and red dashed lines depict the contours with $S_{ij} = S_F$ and $S_{ij} = S_T$, respectively. The resolution component M_R provides limits along the direction of the camera's optical axis while the M_D component generates constraints mainly in the direction across the camera's FOV. In the example shown in Figure 3(a) with $M_R = \alpha f / \hat{z}'$, the handoff margin is only defined at the far end of the camera's FOV along the optical axis. The scenario where the target is so close to the camera that part of it falls out of the camera's FOV is ignored. The modification in (6) imposes a proper constraint at the near end of the camera's FOV along the optical axis. Therefore, the resulting observation is both complete and with desired resolution, as shown in Figure 3(b). Figure 3(c) shows the combined effects of the resolution and distance components, which defines the handoff safety margin.

4. Sensor planning

Based on the definition of OM given in Section 3, our sensor planning algorithm is developed to achieve the optimal balance between coverage and sufficient overlapped margins for successful camera handoff.

Let A_1 represent the grid coverage with $a_{1,ij} = 1$ if $S_{ij} \geq S_F$ and $a_{1,ij} = 0$ otherwise. The A_1 matrix resembles the A matrix in the conventional coverage analysis discussed in Section 2. Two additional matrices are constructed A_2 and A_3. The A_2 matrix has $a_{2,ij} = 1$ if $S_F \leq S_{ij} < S_T$ and $a_{2,ij} = 0$ otherwise. The A_3 matrix has $a_{3,ij} = 1$ if $S_{ij} \geq S_T$ and $a_{3,ij} = 0$ otherwise. Matrices A_2 and A_3 represent the handoff safety margin and visible area, respectively. Let $\mathbf{c'}_k = A_k \mathbf{x}$, $k = 1,2,3$. The objective function at the ith grid is formulated as:

$$c_i = w_1(c'_{1,i} > 0) + w_2(c'_{2,i} = 2) - w_3(c'_{3,i} > 1), \quad (9)$$

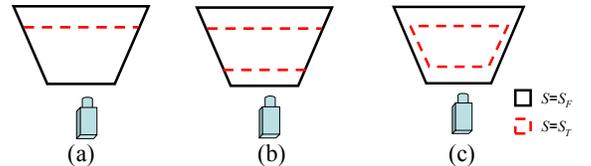where w_1, w_2, and w_3 are predefined positive weights. The

Figure 3. The contours of OM projected onto the ground plane. (a) $S = M_R = \alpha f / \hat{z}'$. (b) $S = M_R$ as defined in (6). (c) $S = w_R M_R + w_D M_D$ with $w_R = 0.5$, $w_D = 0.5$, $\beta_1 = 1$ and $\beta_0 = 1$.

operation $(c'_{1,i} > 0)$ means $(c'_{1,i} > 0) = \begin{cases} 1 & c'_{1,i} > 0 \\ 0 & otherwize \end{cases}$. The

first term in the objective function considers coverage, the second term produces sufficient overlapped handoff margins, and the third term penalizes excessive overlapped visible areas. Our objective function achieves a balance between coverage and adequate margins for camera handoff. The optimal sensor arrangement for the Type 1 and Type 2 problems then can be obtained by:

$$\max \sum_i c_i, \text{ subject to } \sum_j \omega_j x_j \le C_{\max} \qquad (10)$$

$$\min \sum_j \omega_j x_j \text{ and } \max \sum_i c_i, \text{ subject to } \sum_i b_i \ge C_{\min}. \qquad (11)$$

The choice of the weights depends on the importance assigned to coverage and handoff success rate. A rule of thumb is $w_1 < w_2 < w_3$.

4.1. Function validation

In this section, we examine the behavior of the newly defined objective function to validate its effectiveness. For clear illustration, we study the behavior of the objective function based on the relative position between two cameras. The position of one camera is fixed while the position of the other camera has four degrees of freedom including horizontal translation ΔX, vertical translation ΔY, pan θ_P, and tilt θ_T. From the definition of OM, the contours defined by $S_{ij} = S_F$ and $S_{ij} = S_T$ approximately constitute a trapezoid. The corresponding parameters are given in Figure 4.

The derivation of the exact expression of the objective function is not difficult but tedious. To simplify the process and yet reveal the characteristics of the objective function, we fix $\Delta Y = \theta_P = \theta_T = 0$ and study the relation between the objective function and ΔX as our first step. The resulting function $F = \sum_i c_i$ can be expressed as:
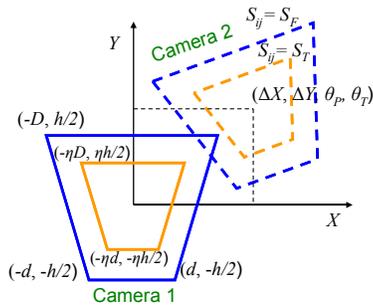

Figure 4. Schematic illustration of the geometric relation between the adjacent cameras' FOVs for computing the objective function. The position of camera 1 is fixed while the position of camera 2 can be varied with four degrees of freedom.

$$\begin{cases} F_1 = 2w_1(D+d)h + (w_2 - w_1)\dfrac{(\Delta X - 2D)^2 h}{4(D-d)} & \Delta X_{th} < \Delta X \le 2D \\[2mm] F_2 = F_1 - w_2 \dfrac{(\Delta X - \Delta X_{th})^2 h}{4(D-d)} & 2\eta D < \Delta X \le \Delta X_{th} \quad (12) \\[2mm] F_3 = F_2 - w_3 \dfrac{(\Delta X - 2\eta D)^2 h}{4(D-d)} & \eta D \le \Delta X \le 2\eta D \end{cases}$$

with $\Delta X_{th} = 2D - (1-\eta)(3D-d)/2$. Since the coverage, overlapped handoff margins, and overlapped visible areas become effective in (9) in sequence as ΔX decreases, the corresponding objective function has three expressions depending on the value of ΔX.

Given the expression in (12), Figure 5 shows the objective function with different choices of weights. We can see that the optimal ΔX^* is achieved with $2\eta D < \Delta X^* < \Delta X_{th}$. When a smaller weight is assigned to the coverage term, the optimal ΔX^* is shifted toward $2\eta D$, resulting in more overlapped FOVs for executing camera handoff. On the contrary, if a larger weight is assigned to the coverage term, the optimal ΔX^* is shifted toward ΔX_{th}, leading to an improved coverage at the cost of decreased overlapped handoff margins. From the derivatives of the functions in (12), we note that F_1 and F_3 are monotonically decreasing and increasing functions, respectively, with a proper choice or the weights in (9). The turning point falls in the range of F_2 and is determined by the relation between w_1 and w_2, the weights for the coverage and handoff margin terms in (9).

4.2. Special performance requirements

For general sensor planning, the objective function defined in (9) can be used. However, when special requirements are to be satisfied, additional constraints need to be included. The resolution consideration corresponds to priority areas which need a specified resolution. The frontal view component points to path constraints where there exist predefined paths within which the objects' movements are restricted.

There exist two approaches to impose the resolution requirements: direct constraint and adaptive weight. To incorporate the resolution requirements, we construct a supplementary matrix A_4 with $a_{4,ij} = 1$ if $M_{R,ij} \ge M_{R,o,i}$ and $a_{4,ij} = 0$ otherwise, where $M_{R,o,i}$ is the corresponding resolution requirement at the ith grid point. The direct constraint approach is carried out by introducing an extra constraint $A_4 \mathbf{x} \ge \mathbf{b}_{R,o}$ where $\mathbf{b}_{R,o}$ represents the required resolution with $b_{R,o,i} = 1$ if the corresponding grid needs the minimum resolution and $b_{R,o,i} = 0$ otherwise. The adaptive weight approach on the other hand assigns different weights $w_{1,i}$ to the grid points according to the coverage requirements. Larger weights are used if the corresponding grids require the minimum resolution. The objective function then becomes:
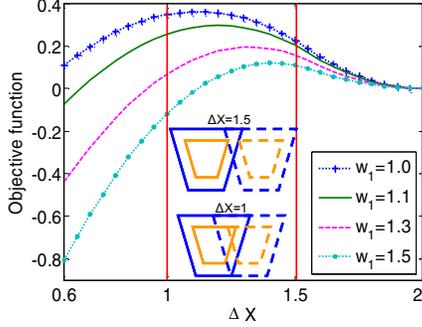
Figure 5. The objective function with varying ΔX and different choices of w_1, the weight assigned to the coverage term in (9). D=1, d=0.6, h=0.8, η=0.6, w_2=2, w_3=5.

$$c_i = w_1(c'_{1,i} > 0) + w_2(c'_{2,i} = 2) \\ - w_3(c'_{3,i} > 1) + w_{4,i}(c'_{4,i} > 0) \quad (13)$$

where $\mathbf{c'}_4 = A_4\mathbf{x}$ and $w_{4,i}$ are different weights allocated according to the resolution requirement.

In surveillance systems, a predefined path is commonly encountered. It is also preferred that a frontal view can be achieved sometime while pedestrians are walking along this path. An example is the entrance areas where a frontal view of the pedestrian is preferred when he or she enters the gate. We use the tangential direction of the middle line of the path as the average direction of the pedestrian's motion. Let the kth point on the middle line be $P_{0,k}$ and its tangential direction $V_{P,k}$. The frontal view measure observed by the jth camera (placed at $T_{C,j}$) at point $P_{i',k}$ along the line perpendicular to $V_{P,k}$ is given by:

$$FV_{i'j} = \frac{\left(T_{C,j} - P_{i',k}\right)^T V_{P,k}}{\| T_{C,j} - P_{i',k} \| \| V_{P,k} \|}. \quad (14)$$

Based on $FV_{i'j}$, we define a matrix A_5 with $a_{5,i'j} = 1$ if $FV_{i'j} \geq 0$ and $a_{5,i'j} = 0$ otherwise. Let $a_{5,ij} = 0$ for grid points outside the path. Finally the frontal view or path constraint is incorporated into the objective function by:

$$c_i = w_1(c'_{1,i} > 0) + w_2(c'_{2,i} = 2) \\ - w_3(c'_{3,i} > 1) + w_{5,i}(c'_{5,i} > 0) \quad (15)$$

where $\mathbf{c'}_5 = A_5\mathbf{x}$. Note that although the resolution and frontal view constraints are addressed separately, it is straightforward to combine these two terms. The only modification is to simply add the corresponding components. The adaptive weight approach is especially attractive because of its concise expression and speed of convergence.

5. Experimental results

In this section, we first examine the soundness of the newly developed OM for perspective cameras and then introduce our experimental methodology. Our experimental results using two typical office floor plans are presented and compared with the reference algorithm proposed by Erdem and Sclaroff [9]. For clear presentation, the reference algorithm is denoted as T1C and T2C for Type1 and Type 2 problems, where C stands for coverage. Our sensor planning methods discussed in Section 4.1 are denoted as T1H and T2H, where H stands for handoff. When the frontal view or path constraint is included, we refer to our methods described in Section 4.2 as T1P and T2P, where P stands for the path constraint. Comparing the T1C (T2C) method with the T1H (T2H) method, an improved handoff success rate is expected. The major difference between the T1H and T1P (T2H and T2P) methods lies in that the frontal view component is added in the T1P and T2P methods. Therefore an improved frontal view percentage is expected from the T1P and T2P methods.

5.1. Experiments on observation measure

A perspective camera is placed at T_C=[0 0 3m]^T looking down towards the ground plane at a tilt angle of -30°. Its pan angle is set to zero. The image size is 640×480. The camera's focal length is 21.0mm. Points are uniformly sampled on the ground plane (Z=0) with X in the range of -8m to 8m and Y in the range of 3m to 10m. Figure 6 shows the corresponding OM values for the perspective camera. The best observation area with the maximum OM values is in the proximity of $[0\ 5m\ 0]^T$. As the object moves away from this area, the OM value decreases. A higher penalty is given to the motion along the X direction, the direction across the camera's FOV. The proposed OM gives a quantified measure of the tracking and observation suitability, which agrees with our intuition and visual inspection. Three disjoint regions are also depicted in the ground plane separated by the failure and trigger thresholds.
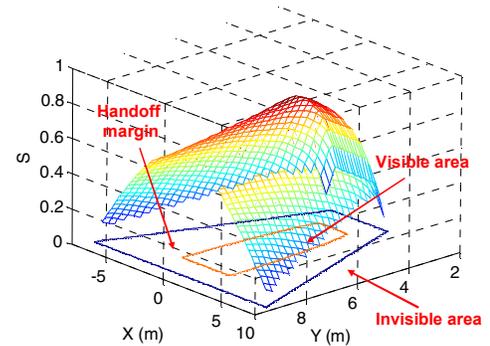

Figure 6. Graphical illustration of the observation measure and handoff safety margin of a perspective camera placed at [0 0 3m].

5.2. Experimental methodology

The floor plans under test are shown in Figures 1(a) and 7. The floor plan in Figure 1(a) represents two types of environments encountered in practical surveillance: space with obstacles (region A illustrated in yellow) and open space where pedestrian can move freely (region B illustrated in green). Region B is deliberately included because it imposes more challenges on camera placement when considering handoff success rate. Camera handoff is relatively easier when there is a predefined path compared with the scenarios where subjects are assumed to move freely, since camera handoff may be triggered at any point in the camera's FOV. Figure 7 illustrates an environment with a predefined path where workers proceed in a predefined sequence. In the following experiments, we will refer to the two floor plans as plan A and B. In our experiments, perspective cameras are free to be placed along the walls of the environment.

To obtain a statistically valid estimation of handoff success rate and frontal view percentage, simulations are carried out to enable a large amount of tests under various conditions. A pedestrian behavior simulator [11, 12] is implemented so that we could have a close resemblance to the experiments in real environments and in turn an accurate estimation of the handoff success rate. To save space, interested readers can refer to the original papers for details. In our experiments, the arrival of the pedestrian follows a Poisson distribution with an average arrival rate of 0.1 (person/second). The average walking speed is 0.5 (meters/second). Several points of interest are generated randomly to form a pedestrian trace. Figures 1(a) and 7 depict some randomly generated pedestrian traces. System performances are evaluated using coverage (C), handoff success rate (HSR), and frontal view percentage (FVP). HSR and FVP are obtained from simulations of 300 randomly generated pedestrian traces.
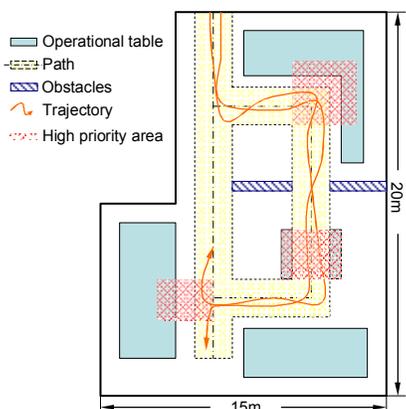
5.3. Experiments on sensor planning

For both floor plans we require a visible distance of 10m and a height of 3m. The same pair of tilt angle and focal length can be used with $f = 21.0mm$ and $\theta_T = -30°$. The importance weights w_1, w_2, and w_3 are set to 1, 2, and 5. $w_5=5$ if the path constraint is imposed. The failure and trigger thresholds are 0 and 0.4, respectively

Figure 8 illustrates the experimental results using floor plan A to solve the Type 1 problem. Our T1H approach chooses a camera positioning scheme with a slightly decreased coverage from 81.6% to 74.7% compared with the T1C method. However, the averaged HSR is improved substantially from 23.2% to 87.4%. An example trace is also shown in Figure 8(c) and (d). As expected, if only coverage is considered, insufficient overlapped areas are kept between the adjacent cameras, leading to two handoff failures as observed in Figure 8(c). In Figure 8(d), the target is tracked continuously with three successful handoffs.

Figure 9 shows and compares sample frames from two cameras with and without sufficient handoff margins. If only coverage is taken into account as shown in Figure 9(a), the object of interest is lost before the left camera is able to identify the subject and cooperate with the right camera. With sufficient handoff margins as shown in Figure 9(b), the object of interest can be detected and labeled correctly before it becomes unidentifiable in the right camera.

As expected, a considerably improved HSR is also achieved for floor plan B, as shown in Figure 10. In addition, we add the frontal view criterion and test the T1P method. The averaged FVP is elevated from 28.7% to 93.5%. From Figures 10(b) and (c), we could see that the cameras are turned towards the direction of the predefined path after introducing the frontal view constraint.
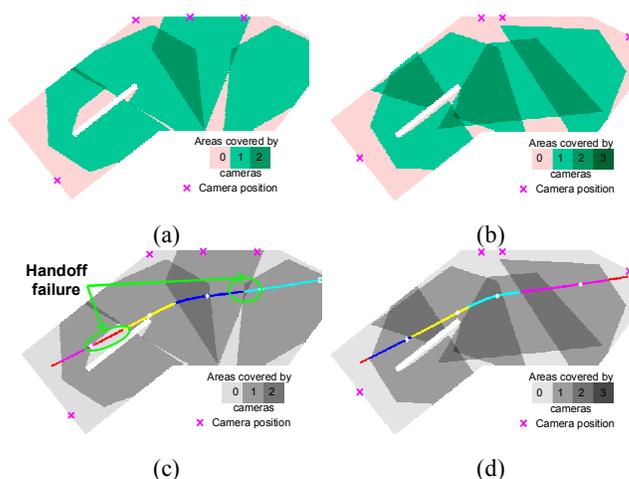

Figure 7. Illustration of an office floor plan with predefined path which the workers' movements are restricted to.

Figure 8. Optimal camera positioning for floor plan A and Type 1 problem (a) T1C (C: 81.6 %, HSR: 23.2%) and (b) T1H (C: 74.7%, HSR: 87.4%). An example trace: two handoff failures in (c) and three successful handoffs in (d).
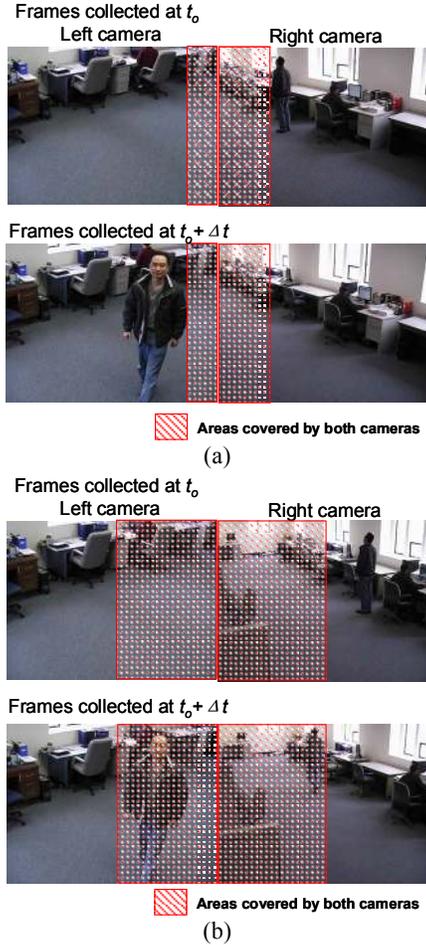
Figure 9. Illustration of sufficient safety margin for automated camera handoff. Sample frames from two cameras (a) when only coverage is considered and (b) when both coverage and handoff success rate are considered. The object of interest is visible in the right camera at t_o. In (a), the object of interest is lost in the right camera as it moves and becomes visible in the left camera at $t_o+\Delta t$. There is no sufficient margin for a successful handoff. In (b), the object of interest remains visible in the right camera at $t_o+\Delta t$, which ensures a successful handoff.

The Type 2 problem imposes additional requirements on the overall coverage, which leaves less freedom in the optimization process to maximize the HSR. As Figure 11 demonstrates, the overall coverage is constrained to be above 80%, which results in a decrease in HSR from 87.4% obtained by the T1H method to 68.5%. However, with a similar coverage (81.5% vs. 81.6%), our T2H algorithm is still able to achieve a much higher HSR (68.5%) than the conventional T1C approach (23.2%). Similar observations apply to the experiments using floor plan B as shown in Figure 12.

5.4. Performance comparison

Table 1 summarizes the performance comparison between our algorithms and the reference algorithm proposed by Erdem and Sclaroff [9]. Consistent observations are obtained from experiments using two floor plans. Compared with the reference algorithm, our algorithms produce considerably improved HSR and FVP at the cost of slightly decreased coverage. This amount of decrease in coverage is inevitable so as to maintain overlapped areas between adjacent cameras required by persistent and automated tracking for a fixed number of cameras.

The ratio between the increase in HSR and the decrease in coverage $\Delta HSR/|\Delta C|$ describes the advantage of our algorithms. For the Type 1 problem, every 1% decrease in
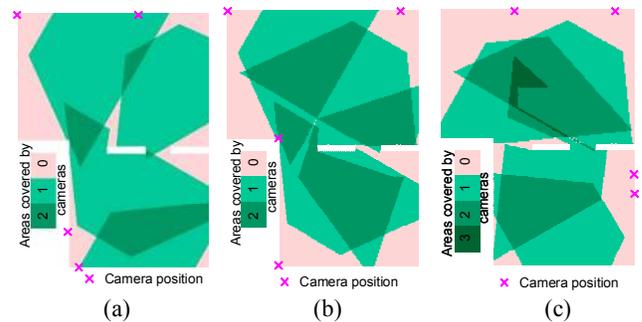

Figure 10. Optimal camera positioning for floor plan B and Type 1 problem: (a) T1C (C: 84.8%, HSR: 6.0%, FVP: 67.7 %), (b) T1H (C: 74.7%, HSR: 56.9 %, FVP: 28.7%), and (c) T1P (C: 72.1%, HSR: 58.0%, FVP: 93.5%).
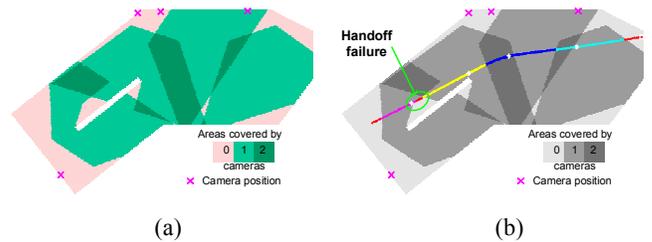

Figure 11. Optimal camera positioning for floor plan A and Type 2 problem (C≥80%). (a) T2H (C: 81.5%, HSR: 68.5%). (b) An example trace with one handoff failure and two successful handoffs.
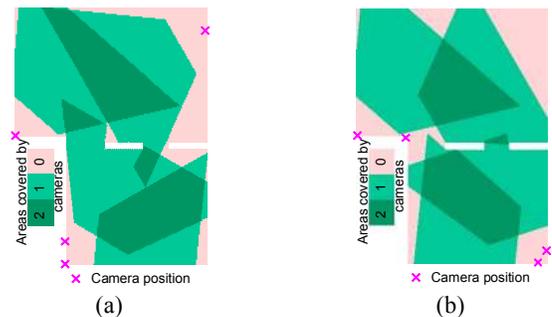

Figure 12. Optimal camera positioning for floor plan B and Type 2 problem (C≥80%): (a) T2H (C: 81.3%, HSR: 43.7%, FVP: 41.0%) and (b) T2P (C: 81.6 %, HSR: 47.1 %, FVP: 69.0%).

Table 1. System performance comparison. C: coverage, HSR: handoff success rate, and FVP: frontal view percentage.

Floor plan A				
Method	C (%)	HSR(%)	ΔHSR/|ΔC|	
T1C/T2C	81.6	23.2		
T1H	74.7	87.4	9.3	
T2H (80%)	81.5	68.5	453	
Floor plan B				
Method	C (%)	HSR(%)	ΔHSR/|ΔC|	FVP(%)
T1C/T2C	84.8	6.0		67.7
T1H	74.7	56.9	5.0	28.7
T1P	72.1	58.0	4.1	93.5
T2H (80%)	81.3	43.7	10.8	41.0
T2P	81.6	47.1	12.8	69.0

coverage results in a 4% to 10% increase in HSR. An even higher improvement rate can be achieved for the Type 2 problem. The efficiency of the proposed algorithms in balancing the overall coverage and sufficient overlapped FOVs becomes evident. Furthermore, our algorithms can handle additional constraints, such as the frontal view requirement. The resulting T1P and T2P algorithms are able to maintain a similar improvement rate in HSR as the general methods (T1H and T2H) with further improved FVP.

Conventional sensor planning methods produce a camera placement with a maximized coverage. In such a system, although a target can be seen, it cannot be automatically and consistently labeled or recognized as the same target across different cameras because of handoff failures resulted from insufficient overlapped FOVs. The resulting camera placement cannot support automated and persistent surveillance systems since the tracked or identified target trajectories are disjointed at the junction areas of adjacent cameras. In contrast, our sensor placement ensures a continuous and consistently labeled trajectory. The decreased coverage can be easily compensated for by adding an additional camera. The cost of an extra camera is acceptable in comparison with a system with inherent disability of maintaining persistent and continuous tracking.

6. Conclusions

In this paper, we proposed a general sensor planning algorithm in the context of persistent and automated tracking and improved existing algorithms by incorporating handoff rate analysis with coverage and visibility analysis. Direct constraint and adaptive weight approaches were derived from the general method to solve the resolution and frontal view constraints. Significantly improved handoff success rate and frontal view percentage were reported via experiments and comparisons with a reference algorithm using typical office floor plans. This indicates the efficiency of the proposed algorithm in balancing the overall coverage and sufficient overlapped FOVs. With considerably improved handoff success rate and frontal view percentage, the proposed algorithm produces robust and enhanced performance compared with the reference algorithm presented in [9] when applied to automated tracking systems.

References

[1] S. D. Roy, S. Chaudhury, and S. Banerjee, "Active recognition through next view planning: a survey", *Pattern Recognition*, vol. 37, no. 3, pp. 429-446, Mar. 2004.

[2] S. D. Roy, S. Chaudhury, and S. Banerjee, "Recognizing large isolated 3-D objects through next view planning using inner camera invariants", *IEEE Trans. on Systems, Man and Cybernetics*, vol. 35, no. 2, pp. 282-292, Apr. 2005.

[3] L. M. Wong, C. Dumont, and M. A. Abidi, "Next best view system in a 3D object modeling task", *IEEE Int'l Symposium on Computational Intelligence in Robotics and Automation*, Monterey, CA, Nov. 1999, pp. 306-311.

[4] S. Yous, N. Ukita, and M. Kidode, "Multiple active camera assignment for high fidelity 3D video", *IEEE Int'l Conf. on Computer Vision Systems*, New York, Jan. 2006.

[5] F. Z. Quereshi and D. Teropoulos, "Towards intelligent camera networks: a virtual vision approach", *IEEE Int'l Workshop on Visual Surveillance and Performance Evaluation of Tracking and Surveillance*, Beijing, China, Oct. 2005, pp. 177-184.

[6] Q. Cai and J. K. Aggarwal. "Tracking human motion in structured environments using a distributed-camera system", *IEEE Trans. on Pattern Recognition and Machine Intelligence*, vol. 21, no. 11, pp. 1241-1247, Nov 1999.

[7] V. Isler, S. Khanna, J. Spletzer, and C.J. Taylor, "Target tracking with distributed sensors: the focus of attention problem", *Computer Vision and Image Understanding*, vol. 100, no. 1-2, Oct.-Nov. 2005, pp. 225-247.

[8] A. Mittal and L. S. Davis, "Visibility analysis and sensor planning in dynamic environments", *European Conf. on Computer Vision*, Prague, Czech Republic, May 2004, pp. 175-189.

[9] U. M. Erdem and S. Sclaroff, "Automated camera layout to satisfy task-specific and floor plan-specific coverage requirements", *Computer Vision and Image Understanding*, vol. 103, no. 3, pp. 156-169, Sept. 2006.

[10] P. J. Phillips, P. Grother, R. J. Micheals, D. M. Blackburn, E. Tabassi, and M. Bone, "Face Recognition Vendor Test 2002, Evaluation Report".

[11] G. Antonini, S. Venegas, M. Bierlaire, and J. Thiran, "Behavioral priors for detection and tracking of pedestrians in video sequences", *Int'l Journal of Computer Vision*, vol. 69, no. 2, pp. 159-180, Aug. 2006.

[12] J. Pettre, T. Simeon, and J. P. Laumond, "Planning human walk in virtual environments", *IEEE/RSJ Int'l Conf. on Intelligent Robots and Systems*, Lausanne, Switzerland, Sept. 2002, pp. 3048-3053.