Chapter 16

Bargaining Strategies for Camera Selection in a Video Network

Bir Bhanu and Yiming Li

Contents

Due to the broad coverage of an environment and the possibility of coordination among different cameras, video sensor networks have attracted much interest in recent years. Although the field of view (FOV) of a single camera is limited and cameras may have overlapping or nonoverlapping FOVs, seamless tracking of moving objects can be achieved by exploiting the handoff capability of multiple cameras. In this chapter, we will provide a new perspective to the camera selection and handoff problem that is based on game theory. In our work, game theory is used for multi-camera multi-person seamless tracking based on a set of user-supplied criteria in a network of video cameras for surveillance and monitoring. The bargaining mechanism is considered for collaborations as well as for resolving conflicts among the available cameras. Camera utilities and person utilities are computed based on a set of criteria. They are used in the process of developing the bargaining mechanisms. The merit of our approach is that it is independent of the topology of how the cameras are placed in the network. When multiple cameras are used for tracking and where multiple cameras can "see" the same object, we are able to choose the "best" camera based on multiple criteria that are selected a priori. The algorithm can automatically provide an optimal as well as stable solution of the camera assignment quickly. The detailed camera calibration or 3D scene understanding is not needed in our approach. Experiments for multi-camera multi-person tracking are provided to corroborate the proposed approach. We also provide a comprehensive comparison of our work and some non-game-theoretic approaches, both theoretically and experimentally.

16.1 Introduction

The growing demand for security in airports, banks, shopping malls, homes, etc. leads to an increasing need for video surveillance, where camera networks play an important role. Significant applications of video network include object tracking, object recognition, and object activities from multiple cameras. The cameras in a network can cooperate with each other and perform various tasks in a collaborative manner. Multiple cameras enable us to have different views of the same object at the same time, such that we can choose one or some of them to monitor a given environment. This can help to solve the occlusion problem to some extent, as long as the FOVs of the cameras have some overlaps. However, since multiple cameras may be involved over long physical distances, we have to deal with the handoff problem as well. *Camera handoff* is the process of finding the next best camera to see the target object when it is leaving the FOV of the current camera, which is being used to track it [17]. This has been an active area of research and many approaches have been proposed. Some camera networks require switches (video matrix) to help monitor the scenes in different cameras [1]. The control can be designed to switch among cameras intelligently. Both distributed and centralized systems are proposed. Some researchers provide hardware architecture design, some of which involve embedded smart cameras, while others focus on the software design for camera assignment and algorithm development. This chapter first gives a comprehensive review for the existing related works and then focuses on an introduction to the game-theoretic approach to do camera selection and handoff, followed by a systematic comparison of this game-theoretic technique with some other non-game-theoretic approaches. Detailed experimental comparisons are provided for four selected techniques.

The rest of this chapter is organized as follows: Section 16.2 gives a comprehensive background of the current and emerging approaches for camera selection and handoff. Comparison tables are provided to help the readers to have a macroscopic view of the existing techniques. Section 16.3 focuses on the theoretical approach description and the comparison with two other non-game-theoretic approaches. Experimental results are provided in Section 16.4. Finally, the conclusions are drawn in Section 16.5.

16.2 Related Work and Our Contributions

The research work in camera selection and handoff for a video network consisting of multiple cameras can be classified according to many different aspects, such as whether it is embedded/PC based; distributed/centralized; calibration needed/calibration free; topology based or topology free; statistics based/statistics free, etc.

16.2.1 Comparison for Existing Works

Some researchers work on the design for embedded smart cameras, which, usually, consist of a video sensor, a DSP or an embedded chip, and a communication module. In these systems, such as [2–6], since all the processing can be done locally, the design work is done in a distributed manner. There are also some PC-based approaches that consider the system in a distributed manner, such as [7–10]. Meanwhile, a lot of centralized systems are proposed as well, such as [11–15]. Some works, such as [15], require the topology of the camera network, while some are image based and do not have requirements for any *a priori* knowledge of the topology. As a result, calibration is needed for some systems, while some systems, such as [16–19] are calibration free. Active cameras (pan/tilt/zoom cameras) are used in some systems, such as [14,15,17], to obtain a better view of objects. However, to our knowledge, only a small amount of work has been done to propose a large-scale active camera network for video surveillance. More large-scale camera networks generally consist of static cameras. Images in 3D are generated in some systems, such as [6]. However, in most approaches proposed for the camera selection and handoff, only 2D images are deployed. There are also other considerations, such as resource allocation [20], fusion of different types of sensors [21], etc. In Table 16.1, we compare the advantages and disadvantages for some of the important issues discussed earlier.

Table 16.2 lists sample approaches from the literature and their properties. It is to be noticed that, not all the distributed systems are realized in an embedded fashion. For instance, a distributed camera node can consist of a camera and a PC as well, although the trend is to realize distributed systems via embedded chips. That is why we treat distributed systems and embedded systems separately in Table 16.1. In Table 16.2, some approaches are tested using real data, while some provide only the simulation results. There is no guarantee that the systems, which are experimented using synthetic data, can still work satisfactorily and realize real-time processing when using real data. So, the real-time property is left blank for those approaches whose experiments use simulated data. Similarly, most of the experiments are done for a small-scale camera network. The performance of the same systems for a large-scale camera network still needs to be evaluated.

Table 16.1 Merits of Various Characteristics Encountered in Distributed Video Sensor Networks

Properties	*Advantages*	*Disadvantages*
Distributed	Low bandwidth requirement; no time requirement for image decoding; easy to increase the number of nodes; the system is hard to die fully	Lack of global cooperation
Centralized	Easy for cooperation among cameras; hardware architecture is relatively simple compared with distributed systems	Require more bandwidth; high computational requirements for the central server; may cause severe problem once the central server is down
Embedded	Easy to realize distributed system; low bandwidth	Limited resources, such as memory, computing performance, and power; only simple algorithms have been used
PC based	Computation can be fast; no specific hardware design requirements, like for embedded chips or DSPs	A bulky solution for many cameras
Calibrated	Can help to know the topology of the camera network; a must for PTZ cameras, if a precise zoom is required	Preprocessing is required; calibration process may be time-consuming
Uncalibrated	No off-line camera calibration is required	Exact topology of cameras difficult
Active cameras	Provide better view of objects; can save the number of cameras by pa/tilt to cover larger monitoring range	Camera calibration may be required, especially when zooming; complex algorithms to account camera motions
Static/mobile cameras	Low cost, high for mobile; easy to determine topology of the camera network; relatively simpler algorithms as compared with those for active (and mobile) cameras	More (statitc) cameras are needed to have a full coverage; have no close-up if the object is not close to any cameras

16.2.2 Our Contributions

The contributions of our work are as follows:

- Game-theoretic approach to do camera selection and handoff is provided. Bargaining mechanism is applied to get to the stable solution.

Table 16.2 A Comparison of Some Properties for Selected Approaches

	HW		*Algorithm/SW*		*Experiment Details*					
Approaches	*E*	*A*	*D*	*C*	*RT*	*RD*	N_C	N_P	*T*	*O*
Quaritsch et al. [3]	Yes	No	Yes	No	Yes	Yes	2	1	Camshift	No
Flech and Straβer [5]	Yes	No	Yes	No	Yes	Yes	1	1	Particle filter	Yes
Park et al. [7]	No	No	Yes	No	N/A	No	20	N/A	N/A	Yes
Morioka et al. [8]	No	No	Yes	No	N/A	No	6	1	N/A	Yes+
Micheloni [9]	No	Yes	Yes	Yes	Yes	Yes	3	3	Kalman filter	Yes
Qureshi and Terzopoulos [10]	No	Yes	Yes	Yes	No	No	16	100	N/A	Yes+
Kattnaker and Zabih [12]	No	No	No	No	Yes	Yes	4	2	Bayesian	No
Everts et al. [14]	No	Yes	No	Yes	Yes	Yes	1	1	Histogram-based	No
Javed et al. [16]	No	No	No	No	Yes	Yes	2	2	N/A	Yes
Jo and Han [19]	No	No	No	No	Yes	Yes	2	N/A	Manual	Yes
Gupta et al. [22]	No	No	No	No	Yes	Yes	15	5	M2Tracker	Yes
Song et al. [23]	No	No	Yes	Yes	No	Yes	7	9	Particle filter	No
Song et al. [24]	No	Yes	Yes	Yes	No	No	14	N/A	N/A	Yes
Our approach	No	No	No	No	Yes	Yes	3	2	Particle filter	Yes+

Legends for the table: E—Embedded; A—Active camera; D—Distributed; C—Calibration needed; RT—Real-time; RD—Real data; N_C—Number of cameras; N_P—Number of objects; T—Tracking algorithm used; O—Overlapping FOVs, Yes+—Yes but not necessary.

- A comprehensive comparison of recent work for camera selection and handoff is provided. Two non-game-theoretic approaches are compared with our work both theoretically and experimentally.
- Results with real data and simulations in various scenarios are provided for an in-depth understanding of the advantages and weaknesses of the key approaches. The focus of comparison is solely on multi-object tracking using non-active multi-cameras in an uncalibrated system. The comparison considers software- and algorithm-related issues. Resource allocation, communication errors, and hardware considerations are not considered.

16.3 Technical Approach

16.3.1 Motivation and Problem Formulation

Game theory can be used for analyzing the interactions as well as conflicts among multiple agents [25,26]. Analogously, in a video sensor network, communications as well as competitions among cameras exist simultaneously. The cooperation lies in the fact that all the available cameras, those which can "see" the target person, have to work together to track the person so that he can be followed as long as possible. On the other hand, the available cameras also compete with each other for the rights of tracking this person, so that a camera can maximize its own *utility*. This enlightens us to view the camera assignment problem in a game-theoretic manner. The interactive process is called a game [27], while all the participants of the game are called players, who strive to maximize their *utilities*. The utility of a player refers to the welfare that he can get in the game. In our problem, for each person to be tracked, there exists a multiplayer game, with the available cameras being the players. If there are multiple persons in the system, this becomes a multiple of multiplayer game being played simultaneously.

Vehicle–target assignment [28] is a multiplayer game that aims to allocate a set of vehicles to a group of targets and achieve an optimal assignment. Viewing the persons being tracked as "vehicles" while the cameras as "targets," we can adopt the vehicle–target assignment model to choose the "best" camera for each person. In the following, we propose a game theory–based approach that is well suited to the task at hand.

16.3.2 Game-Theoretic Framework

Game theory involves *utility*, the amount of "welfare" an agent derives in a game. We are concerned with three different utilities:

1. *Global utility*: the overall degree of satisfaction for tracking performance.
2. *Camera utility*: how well a camera is tracking persons assigned to it based on the user-supplied criteria.
3. *Person utility*: how well a person is satisfied while being tracked by some camera.

Our objective is to maximize the global utility as well as to make sure that each person is tracked by the "best" camera. During the course of competition among available cameras, they *bargain* with each other, and finally a decision is made for the best camera assignment based on a set of probabilities.

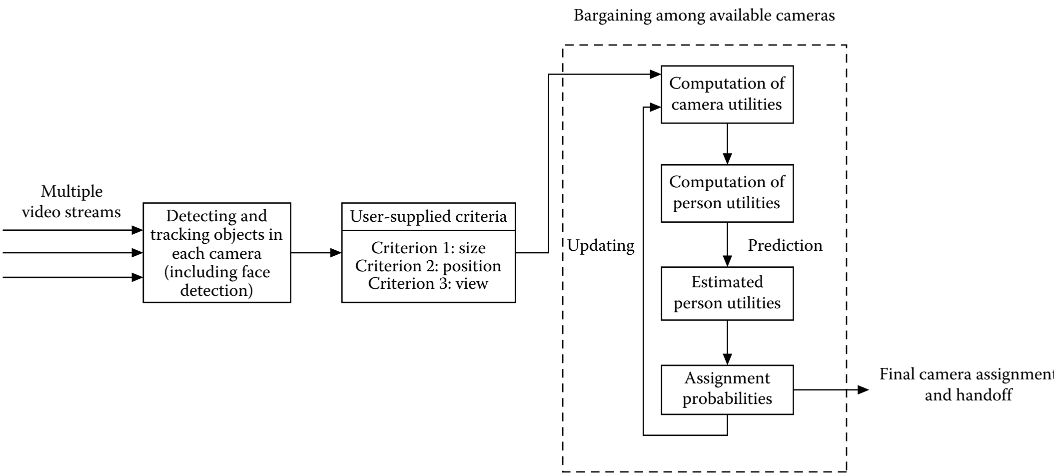

Figure 16.1 Game-theoretic framework for camera assignment and handoff.

An overview of the approach is illustrated in Figure 16.1. Moving objects are detected in multiple video streams. Their properties, such as the size of the minimum bounding rectangle and other region properties (color, shape, location within FOV, etc.), are computed. Various utilities (camera utility, person utility, and global utility) are computed based on the user-supplied criteria, and bargaining processes among available cameras are executed based on the prediction of person utilities at each step. The results obtained from the strategy execution are in turn used for updating the camera utilities and the person utilities until the strategies converge. Finally, those cameras with the highest converged probabilities are used for tracking. This assignment of persons to the "best" cameras leads to the solution of the handoff problem in multiple video streams.

A set of symbols are used in the discussion of our approach and their descriptions are given in Table 16.3.

16.3.2.1 Computation of Utilities

We first define the following properties of our system:

1. A person P_i can be in the FOV of more than one camera. The available cameras for P_i belong to the set A_i. C_0 is assumed to be a virtual (null) camera.
2. A person can only be assigned to one camera. The assigned camera for P_i is named as a_i.
3. Each camera can be used for tracking multiple persons.

For some person P_i, when we change its camera assignment from a' to a'' while assignments for other persons remain the same, if

$$U_{P_i}\left(a'_i, a_{-i}\right) < U_{P_i}\left(a''_i, a_{-i}\right) \Leftrightarrow U_g\left(a'_i, a_{-i}\right) < U_g\left(a''_i, a_{-i}\right) \tag{16.1}$$

Table 16.3 Notations of Symbols Used in the Paper

Symbols	*Notations*
P_{i}	Person *i*
C_j	Camera *j*
N_p	Total number of persons in the entire network at a given time
N_c	Total number of cameras in the entire network at a given time
A_i	The set of cameras that can see person *i*, $A_i = \{a_1, a_2, \ldots, a_{n_C}\}$
n_c	Number of cameras that can *see object i, number of elements in* A_i
n_{p}	Number of persons currently assigned to camera C_j
a_{i}	The assigned "best" camera for person *i*
$a_{-\mathrm{i}}$	The assignment of cameras for the persons excluding person *i*
a	Assignment of cameras for all persons, $a = (a_{\mathrm{i}},\ a_{-\mathrm{i}})$
$U_{C_j}(a)$	Camera utility for camera *j*
$U_{P_i}(a)$	Person utility for person *i*
$U_g(a)$	Global utility
$\bar{U}_{P_i}(k)$	Predicted person utility for person *i* at step *k*, $\bar{U}_{P_i}(k) = \left[\bar{U}^1_{P_i}(k), \ldots, \bar{U}^l_{P_i}(k), \ldots, \bar{U}^{n_c}_{P_i}(k)\right]$, where $\bar{U}^l_{P_i}(k)$ is the predicted person utility for P_{i} if camera a_{i} is used
$p_{\mathrm{i}}(k)$	Probability of person *i*'s assignment at step *k*, $p_i(k) = \left[p_i^1(k), \ldots, p_i^l(k), \ldots, p_i^{N_C}(k)\right]$, where $p_i^l(k)$ is the probability for camera a_{i} to track person P_{i}

the person utility U_{P_i} is said to be aligned with the global utility U_g, where a_{-i} stands for the assignments for persons other than P_i, i.e., $a_{-i} = (a_1, \ldots, a_{i-1}, a_{i+1}, \ldots, a_{N_P})$. We define the global utility as

$$U_g(a) = \sum_{C_j \in C} U_{C_j}(a) \tag{16.2}$$

where $U_{C_j}(a)$ is the camera utility and defined to be the utility generated by all the engagements of persons with a particular camera C_j. Now, we define the person utility as

$$U_{P_i}(a) = U_g(a_i, a_{-i}) - U_g(C_0, a_{-i}) = U_{C_j}(a_i, a_{-i}) - U_{C_j}(C_0, a_{-i}) \tag{16.3}$$

The person utility $U_{P_i}(a)$ can be viewed as a marginal contribution of P_i to the global utility. To calculate (16.3), we have to construct a scheme to calculate the camera utility $U_{C_j}(a)$. We assume that there are N_{Crt} criteria to evaluate the quality of a camera used for tracking an object. Thus, the camera utility can be built as

$$U_{C_j}(a_i, a_{-i}) = \sum_{s=1}^{np} \sum_{l=1}^{N_{Crt}} Crt_{sl} \tag{16.4}$$

where n_p is the number of persons that are currently assigned to camera C_j for tracking. Plugging (16.4) into (16.3) we can obtain

$$U_{P_j}(a_i, a_{-i}) = \sum_{s=1}^{np}\sum_{l=1}^{N_{Crt}} Crt_{sl} - \sum_{\substack{s=1 \\ s=P_i}}^{np}\ \sum_{l=1}^{N_{Crt}} Crt_{sl} \tag{16.5}$$

where s $\neq P_i$ means that we exclude person P_i from those who are being tracked by camera C_j. One thing to be noticed here is that when designing the criteria, we have to normalize them.

16.3.2.2 Bargaining among Cameras

As stated previously, our goal is to optimize each person utility as well as the global utility. Competition among cameras finally leads to the Nash equilibrium. Unfortunately, this Nash equilibrium may not be unique. Some of them are not stable solutions, which are not desired. To solve this problem, a bargaining mechanism among cameras is introduced, to make them finally come to a compromise and generate a stable solution.

When bargaining, the assignment in the kth step is made according to a set of probabilities

$$p_i(k) = \left[p_i^1(k), \ldots, p_i^l(k), \ldots, p_i^{nC}(k)\right]$$

where n_c is the number of cameras that can "see" the person P_i and $\sum_1^{nC} p_i^l(k) = 1$, with each $0 \leq p_i^l(k) \leq 1,\ l = 1, \ldots, n_C$. We can generalize $p_i(k)$ to be

$$p_i(k) = \left[p_i^1(k), \ldots, p_i^l(k), \ldots, p_i^{N_C}(k)\right]$$

by assigning a zero probability for those cameras that cannot "see" the person P_i, meaning that those cameras will not be assigned according to their probability. Thus, we can construct an $N_p \times N_C$ probability matrix

$$\begin{bmatrix} p_1^1(k) & \cdots & p_1^{N_C}(k) \\ \vdots & \ddots & \vdots \\ p_{N_P}^1(k) & \cdots & p_{N_P}^{N_C}(k) \end{bmatrix}$$

At each bargaining step, we will assign a person to the camera that has the highest probability. Since in most cases, a person has no information of the assignment before it is made, we introduce the concept of predicted person utility $\bar{U}_{P_i}(k)$: Before we decide the final assignment profile, we predict the person utility using the previous person's utility information in the bargaining steps. As shown in (16.5), person utility depends on the camera utility, so we predict the person utility for every possible camera that may be assigned to track it. Each element in $\bar{U}_{P_i}(k)$ is calculated by the following equation:

$$\bar{U}_{P_i}^l(k+1) = \begin{cases} \bar{U}_{P_i}^l(k) + \dfrac{1}{p_i^l(k)}(U_{P_i}(a(k)) - \bar{U}_{P_i}^l(k), a_i(k) = A_i^l \\ \bar{U}_{P_i}^l(k), \quad otherwise \end{cases} \tag{16.6}$$

with the initial state $\bar{U}^l_{P_i}(1)$ to be assigned arbitrarily as long as it is within the reasonable range for $\bar{U}_{P_i}(k)$, for $l = 1, \ldots, n_C$. Once these predicted person utilities are calculated, it can be proved that the equilibrium for the strategies lies in the probability distribution that maximizes its perturbed predicted utility [10],

$$P_i(k)' \bar{U}_{P_i}(k) + \tau H(p_i(k)) \tag{16.7}$$

where

$$H(p^l_i(k)) = -p^l_i(k)' \log(p^l_i(k)) \tag{16.8}$$

is the entropy function and τ is a positive parameter belonging to [0,1] that controls the extent of randomization. The larger the τ is, the faster the bargaining process converges; the smaller the τ is, the more accurate result we can get. So, there is a trade-off when selecting the value of τ, and we select τ to be 0.5 in our experiments. The solution of (16.7) is proved [28] to be

$$p^l_i(k) = \frac{e^{((1/\tau)\bar{U}^l_{P_i}(k))}}{e^{((1/\tau)\bar{U}^l_{P_i}(k))} + \cdots + e^{((1/\tau)\bar{U}^{rC}_{P_i}(k))}} \tag{16.9}$$

After several steps of calculation, the result of $p_i(k)$ tends to converge. Thus, we finally get the stable solution, which is proved to be at least suboptimal [28].

16.3.2.3 Criteria for Camera Assignment and Handoff

A number of criteria, including human biometrics, can be used for camera assignment and handoff. For easier comparison between the computed results and the intuitive judgment, four criteria are used for a camera selection:

1. The **size** of the tracked person. It is measured by the ratio of the number of pixels inside the bounding box of the person to that of the size of the image plane. Here, we assume that neither a too-large nor a too-small object is convenient for observation. Assume that λ is the threshold for best observation, i.e., when $r = \lambda$ this criterion reaches its peak value, where

$$r = \frac{\text{\# of pixels inside the bounding box}}{\text{\# of pixels in the image plane}}.$$

$$Crt_{i1} = \begin{cases} \dfrac{1}{\lambda} r, & \text{when } r < \lambda \\ \dfrac{1 - r}{1 - \lambda}, & \text{when } r \geq \lambda \end{cases} \tag{16.10}$$

2. The **position** of the person in the FOV of a camera. It is measured by the Euclidean distance that a person is away from the center of the image plane

$$Crt_{i2} = \frac{\sqrt{(x - x_C)^2 + (y - y_C)^2}}{\left(\frac{1}{2}\right)\sqrt{x_C^2 + y_C^2}} \tag{16.11}$$

where

(x, y) is the current position of the person

(x_c, y_c) is the center of the image plane

3. The **view** of the person, as measured by the ratio of the number of pixels on the detected face to that of the whole bounding box, which is similar to Criterion 1. We assume that the threshold for best frontal view is R, i.e., when $R = \xi$ the view of the person is the best, where

$$R = \frac{\text{\# of pixels on the face}}{\text{\# of pixels on the entire body}}$$

$$Crt_{i3} = \begin{cases} \dfrac{1}{\xi} r, & \text{when } R < \xi \\ \dfrac{1-R}{1-\xi}, & \text{when } R \geq \xi \end{cases} \tag{16.12}$$

4. **Combination** of criteria (1) through (3), which is called the *combined criterion*, is given by the following equation:

$$Crt_{i4} = \sum_{m=1}^{3} w_m Crt_{im} \tag{16.13}$$

where w_m is the weight for different criteria.

It is to be noticed that all these criteria are normalized for calculating the corresponding camera utilities.

16.3.3 Theoretical Comparison with Two Non-Game-Theoretic Approaches

We selected four approaches [8,19] for comparison. They are chosen as typical approaches because these approaches cover both distributed system [8] and centralized system [19]. Although none of these approaches needs camera calibration, some of them do a geometry correspondence [19], while some do not [8]. This section focuses on the comparison of theoretical ideas while experimental comparison is provided in the next section.

In this section, we first describe the key ideas of these approaches. Analysis of the advantages and disadvantages are provided in Table 16.4.

16.3.3.1 Descriptions of the Key Ideas of Selected Approaches

16.3.3.1.1 Approach 1: The Co-Ocurrence to Occurrence Ratio Approach

This approach decides whether two points are in correspondence with each other by calculating the co-occurrence to occurrence ratio (COR). If the COR is higher than some predefined threshold, then the two points are decided to be in correspondence with each other. When one point is getting close to the edge of the FOV of one camera, the system will hand off to another camera that has its corresponding point.

Table 16.4 Relative Merits of the Selected Approaches

Approaches	*Pros*	*Cons*
COR approach [19]	Intuitive efficient approach; acceptable results when there are few occlusions and few cameras and objects	Time-consuming correspondence of point pairs; when correspondence fails or occlusion happens, there is handoff ambiguity and the error rate increases; computing structure becomes complicated with the increase of the number of camera nodes/objects; FOVs have to be overlapped
Fuzzy-based approach [8]	Distributed approach; camera state transition and handoff rules are both intuitive; no requirement for overlapping FOVs	Only simulation results are provided; tracking has to be accurate; not robust when occlusion happens; no guarantee for convergence in a large-scale network
Our approach	Provides a mathematical framework; can deal with the cooperation and competition among cameras; can perform camera selection based on user-supplied criteria; no need for overlapping FOVs	Communication among cameras is not involved; can be extended for distributed computation

The COR is defined as

$$R(x, x\prime) = \frac{p(x, x\prime)}{p(x)} \tag{16.14}$$

where

$$p(x) = \frac{1}{T}\sum_{t=1}^{T}\sum_{i=1}^{N_t} K_2\left(x - x_t^i\right) \tag{16.15}$$

is the mean probability that a moving object appears at x, i.e., the occurrence at x. K_2 is claimed to be circular Gaussian kernel. Similarly,

$$p(x, x\prime) = \frac{1}{T}\sum_{t=1}^{T}\sum_{i=1}^{N_t} K_2\left(x - x_t^i\right)\sum_{i=1}^{N_t'} K_2\left(x\prime - x\prime_t^i\right) \tag{16.16}$$

is the co-occurrence at x in one camera and x' in another camera.

It is intuitive that if two points x and x' are in correspondence, i.e., the same point in the views of different cameras, then the calculated COR should be 1 ideally. On the contrary, if the x and

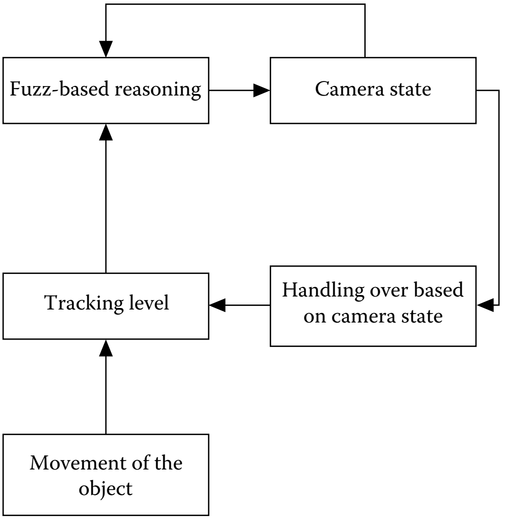

Figure 16.2 Diagram for camera state transition.

x' are completely independent of each other, i.e. two distinctive points, then $p(x, x) = p(x)p(x')$, which leads the COR $R(x, x')$ to be $p(x')$. These are the two extreme cases. If we chose some threshold θ_r such that $p(x') < \theta_r < 1$, then by comparing with θ_r, the correspondence of two points in two camera views can be determined. Another threshold θ_0 is needed to be compared with $p(x)$ to decide whether a point is detected in a camera. Thus, camera handoff can be taken care of by calculating the correspondence of pairs of points in the views of different cameras and performed when necessary.

16.3.3.1.2 Approach 2: Fuzzy-Based Approach

This is another decentralized approach. Each candidate camera has two states for the object that is in its FOV: the **nonselected** state and the **selected** state for tracking. Then, camera handoff is done based on the camera's previous state S_i and the tracking level state SS_i, which is defined by estimating the position measurement error in the monitoring area. The two states for the tracking level are **unacceptable**, meaning that the object is too far away, and **acceptable**, meaning that the object is within the FOV and the quality is acceptable.

The block diagram for camera state transition and the fuzzy rule for camera handoff are given in Figures 16.2 [8] and 16.3 [8], respectively.

16.3.3.2 Pros and Cons Comparison of the Selected Approaches

We list the Pros and Cons for these approaches in Table 16.4.

16.4 Experimental Results

In this section, we perform experiments in different cases for the proposed approach and compare it with the other two non-game-theoretic approaches introduced previously in Section 16.3.3.

(1) If S_i = ***Selected*** And SS_i = ***Acceptable*** Then S_i = ***Selected***
(2) If S_i = ***Non-selected*** And SS_i = ***Unacceptable*** Then S_i = ***Non-selected***
(3) If S_i = ***Selected*** And SS_i = ***Non-selected*** And SS_k = ***Unacceptable*** Then S_i = ***Selected***, $\forall k \in [1, N]$, $k \neq i$, where N is the number of camera candidates
(4) If S_i = ***Non-selected*** And SS_i = ***Acceptable*** And S_k = ***Non-selected*** And SS_k = ***Unacceptable*** Then S_i = ***Selected***, $\forall k \in [1, N]$, $k \neq i$
(5) If S_i = ***Non-selected*** And SS_i = ***Acceptable*** And S_k = ***Selected*** And SS_k = ***Acceptable*** Then S_i = ***Non-selected***, $\equiv k \in [1, N]$, $k \neq i$
(6) If S_i = ***Selected*** And SS_i = ***Unacceptable*** And S_k = ***Non-selected*** And SS_k = ***Acceptable*** Then S_i = ***Non-selected***, $\equiv k \in [1, N]$, $k \neq i$
(7) If S_i = ***Non-selected*** And SS_i = ***Acceptable*** And S_k = ***Selected*** And SS_k = ***Unacceptable*** Then S_i = ***Selected***, $\equiv k \in [1, N]$, $k \neq i$

Figure 16.3 Fuzzy-based reasoning rules.

Although some of the approaches [8] do not have results with real data, in this chapter, both indoor and outdoor experiments with real data are carried out for all the approaches. For convenience of comparison among different approaches, no cameras are actively controlled.

16.4.1 Data

The experiments are done using commercially available AXIS 215 cameras. Three experiments are carried out with an increase in complexity. *Case 1:* two cameras and three persons, indoor. *Case 2*: three cameras and five persons, indoor. *Case 3:* four cameras and six persons, outdoor. The frames are dropped whenever the image information is lost during the transmission. The indoor experiments use cable-connected cameras, with a frame rate of 30 fps. However, for the outdoor experiment, the network is wireless. Due to the low quality of the images, the frame rate is only 10–15 fps on average. The images are 60% compressed for the outdoor experiment to save bandwidth. Image quality is 4CIF, which is 704 × 480. They are overlapped randomly in our experiments, which is not required by some of the approaches but required by some others.

16.4.2 Tracking

None of the approaches discussed here depends on any particular tracker. Basically, ideal tracking can be assumed for comparing the camera selection and handoff mechanisms. It should be noted that tracking is not the focus of our work.

Trackings in all the experiments are initialized by a human observer manually at the very beginning and then done with color-based particle filter [29] automatically. The dynamic model used is random walk. Measurement space is two dimensional: hue and saturation values of a pixel. The sample number used for each object to be tracked is 200 for indoor experiments and 500 for outdoor experiments. Tracking can be done in real time by implementing the OpenCV structure CvConDensation and the corresponding OpenCV functions. Matches for objects are done by calculating the correlation of the hue values using cvComparehist. (We compare the hue values of the upper bodies first. If there is ambiguity, then lower body is considered.) Minor occlusion is recoverable within a very short time. Tracking may fail when severe occlusion takes place or in the case where an object is not in the scene for too long and then reenters. Theoretically, this can be solved by spreading more particles. However, more particles may be very computationally expensive. Thus, we just re-initialize the tracking process manually to avoid non-real-time processing.

16.4.3 Parameters

We first define the following properties of our system:

- A person P_i can be in the FOV of more than one camera. The available cameras for P_i belong to the set A_i.
- A person can only be assigned to one camera. The assigned camera for P_i is named as a_i.
- Each camera can be used for tracking multiple persons.

1. *Our Approach:* In our experiments, we use the combined criterion to perform the bargaining mechanism. This is because it can comprehensively consider all the criteria provided by the user. We give value to the parameters empirically. $\lambda = \frac{1}{15}$, $\xi = \frac{1}{6}$, $w_1 = 0.2$, $w_2 = 0.1$, $w_3 = 0.7$. For instance, the criterion for P_i is calculated as

$$Crt_i = 0.2Crt_{i1} + 0.1Crt_{i2} + 0.7Crt_{i3} \tag{16.17}$$

   The weights are like this because we want to have the frontal view of a person, which contains much information, whenever it is available. The utility functions are kept exactly the same as stated in Section 16.3.2.3.
2. *The COR Approach:* The COR approach in [19] has been applied to two cameras only. We generalize this approach to the cases with more cameras by comparing the accumulated COR in the FOVs of multiple cameras. We randomly select 100 points on the detected person, train the system for 10 frames to construct the correspondence for these 100 points, calculate the cumulative CORs in the FOVs of different cameras, and select the one with the highest value for handoff.
3. *Fuzzy-Based Approach:* We apply the same fuzzy reasoning rule as the one in Figure 16.2, which is given in [8]. The tracking level state is decided by Criterion 2, i.e., Crt_{i2}, which is used for the utility-based game-theoretic approach.

16.4.4 Experimental Results and Analysis

Due to limited space, only those frames with camera handoffs are shown (actually, only some typical handoffs, since the video is long and there are too many handoffs.). These camera handoffs for Cases 1–3 are shown in Figures 16.4 through 16.6, respectively. Since no topology of the camera network is given, tracking is actually performed by every camera all the time. However, for easy observation, we only draw the bounding box for an object in the image of the camera that is selected to track this object. Cases 1 and 2 are simple in the sense that there are fewer cameras and objects, and the frame rate is high enough to make the objects trajectories continuous. So, we only show some typical frames for these cases and give more handoff examples in Case 3, which is more complicated. We show some typical handoffs for Cases 1 and 3, while for Case 2, we show the same frames for the four approaches to see the differences caused by performing handoffs by different approaches.

Figure 16.4 Selective camera handoff frames for the four approaches (Case 1).

Figure 16.5 Selective camera handoff frames for the four approaches (Case 2).

It is clear that the proposed game-theoretic approach considers more criteria when performing the camera selection. Camera handoffs take place whenever a better camera is found based on the user-supplied criterion in this case. So, cameras that can see persons' frontal views, which have the highest weight in Crt_i, are more preferred most of the time. The other two approaches have similar results in the sense that they all consider handoff based on the position of the objects. In this sense, the game-theoretic approach is more flexible to perform camera handoffs based on different criteria. The modification of a criterion will have no influence on the decision-making mechanism.

Figure 16.6 Selective camera handoff frames for the four approaches in Case 3.

Figure 16.4 shows the camera handoff results for a very simple case. All the three approaches achieve similar results, although the game-theoretic approach prefers frontal view.

CAs the scenario being more complex, i.e., more objects and more cameras are involved and occlusions happen frequently, the COR approach and the fuzzy-based approach have less satisfactory results. Error rates for different approaches in each case are given in Table 16.5.

Table 16.5 Error Rates of the Selected Approaches

	Our Approach (%)	*COR (%)*	*Fuzzy Based (%)*
Case 1	***3.86***	***4.23***	***4.64***
Case 2	***4.98***	***10.01***	***7.11***
Case 3	7.89	45.67	21.33

16.5 Conclusions and Future Work

In this work, we propose the novel idea of doing the camera selection and handoff problem in a game-theoretic manner. Some intuitive criteria are designed for easy observation. The bargaining mechanism in game theory is applied to obtain a stable as well as optimal solution to the problem. We also compare our work with two selected non-game-theoretic approaches, which are discussed in detail. Experimental results are provided to show the merits of the proposed game-theoretic approach. It is obvious, from the shown results, that our approach is flexible to deal with multiple predefined criteria. Meanwhile, this provides a systematic method to solve the camera selection problem. As the complexity of the scenario goes up, or the criteria are changed, we do not bother to modify the algorithm and can still have acceptable results.

We also analyzed existing and emerging techniques for the camera selection and handoff problem in the related work part. Advantages and disadvantages of some properties, such as distributed or centralized systems, are discussed.

There is the trend to have a hierarchical structure, which hybrids the distributed and centralized control. In our future work, we will allow communications among cameras to make the algorithm decentralized. Also, there is a lack of research on camera selection and handoff in a large-scale network of active cameras. Current research is short on experimental results with real data processed in real time. We also want to extend our work in a large-scale camera network and realize real-time control of the cameras.

References

1. M. Valera and S.A. Velastin, Intelligent distributed surveillance systems: A review, *IEEE Proc. Vis. Image Signal Process.*, 152, 2, April 2005, 192–204.
2. H. Kim, C. Nam, K. Ha, O. Ayurzana, and J. Kwon, An algorithm of real time image tracking system using a camera with pan/tilt motors on an embedded system, ICMIT, *Proceedings of SPIE*, Vol. 6041 604112–1, 2005.
3. M. Quaritsch, M. Kreuzthaler, B. Rinner, H. Bischof, and B. Strobl, Autonomous multicamera tracking on embedded smart cameras, *EURASIP J. Embedded Systems*, Vol. 2007, 1, January 2007.
4. B. Rinner, M. Jovanovic, and M. Quaritsch, Embedded middleware on distributed smart cameras, *ICASSP 2007*, Honolulu, HI, pp. IV-1381–1384, 2007.
5. S. Flech and W. Straβer, Adaptive probabilistic tracking embedded in a smart camera, *CVPR 2005*, San Diego, CA, 2005.
6. S. Flech, F. Busch, P. Biber, and W. Straβer, 3D surveillance—A distributed network of smart cameras for real-time tracking and its visualization in 3D, *CVPRW 2006*, New York.
7. J. Park, P.C. Bhat, and A.C. Kak, A look-up table based appraoch for solving the camera selection problem in large camera networks, *Workshop on DSC 2006*, Boulder, CO, 2006.

8. K. Morioka, S. Kovacs, J. Lee, P. Korondi, and H. Hashimoto, Fuzzy-based camera selection for object tracking in a multi-camera system, *IEEE Conference on Human System Interactions*, Krakow, Poland, pp. 767–772, 2008.
9. C. Micheloni, G.L. Foresti, and L. Snidaro, A network of co-operative cameras for visual surveillance, *IEE Proc. Vis. Image Signal Process.*, 152, 2, April 2005, 205–212.
10. F. Bashir and F. Porikli, Multi-camera control through constraint satisfaction for persistent surveillance, *AVSS*, Santa Fe, NM, pp. 211–218, 2008.
11. S. Lim, L.S. Davis, and A. Elgannal, A scalable image-based multi-camera visual surveillance system, *AVSS*, Miami, FL, 2003.
12. V. Kattnaker and R. Zabih, Bayesian multi-camera surveillance, *CVPR*, Vol. 2, pp. 253–259, 1999.
13. Y. Lu and S. Payandeh, Cooperative hybrid multi-camera tracking for people surveillance, *Can. J. Elect. Comput. Eng.*, 33, 3/3, Summer/Fall, 145–152, 2008.
14. I. Everts, N. Sebe, and G. Jones, Cooperative object tracking with multiple PTZ cameras, *ICIAP*, Modena, Italy, pp. 323–330, 2007.
15. S. Kang, J. Paik, A. Koschan, B. Abidi, and M.A. Abidi, Real-time video tracking using PTZ cameras, *Proceedings of SPIE*, Getlinburg, Tennessee, Vol. 5132, pp. 103–112.
16. O. Javed, S. Khan, Z. Rasheed, and M. Shah, Camera handoff: Tracking in multiple uncalibrated stationary cameras, *IEEE Workshop on Human Motion*, Austin, TX, pp. 113–118, 2000.
17. U.M. Erdem and S. Sclaroff, Look there! Predicting where to look for motion in an active camera network, *AVSS*, Como, Italy, pp. 105–110, 2005.
18. B. Moller, T. Plotz, and G.A. Fink, Calibration-free camera hand-over for fast and reliable person tracking in multi-camera setup, *ICPR 2008*, Tampa, FL, 2008.
19. Y. Jo and J. Han, A new approach to camera hand-off without camera calibration for the general scene with non-planar ground, *VSSN'06*, Santa Barbara, CA.
20. C. Chen, Y. Yao, D. Page, B. Abidi, A. Koschan, and M. Abidi, Camera handoff with adaptive resource management for multi-Camera multi-target surveillance, *AVSS 2008*, Santa Fe, NM, 2008.
21. J. Nayak, L. Gonzalez-Argueta, B. Song, A. Roy-Chowdhury, and E. Tuncel, Multi-target tracking through opportunistic camera control in a resource constrained multimodal sensor network, *ICDSC 2008*, Stanford, CA, 2008.
22. A. Gupta, A. Mittal, and L.S. Davis, COST: An approach for camera selection and multi-object inference ordering in dynamic scenes, *ICCV 2007*, Rio de Janeiro, Brazil, pp. 1–8.
23. B. Song and A.K. Roy-Chowdhury, Robust tracking in a camera network: A multi-objective optimization framework, *IEEE J. Selected Topics Signal Process.*, 2, 4, Aug. 2008.
24. B. Song, C. Soto, A.K. Roy-Chowdhury, and J.A. Farrell, Decentralized camera network control using game theory, *ICDSC Workshop 2008*, Stanford, CA, 2008.
25. M. Isard and A. Blake, CONDENSATION—Conditional density propagation for visual tracking, *IJCV 1998*, 29(1), pp. 5–28.
26. R.B. Myerson, *Game Theory—Analysis of Conflict*. Harvard University Press, Cambridge, MA, 1991.
27. M.J. Osborne, *An Introduction to Game Theory*. Oxford University Press, New York, 2003.
28. Stanford Encyclopedia of Philosophy, http://plato.stanford.edu/entries/game-theory/#Uti
29. G. Arslan, J.R. Marden, and J.S. Shamma. Autonomous vehicle-target assignment: A game-theoretical formulation. *ASME J. Dynamic Systems Measurement Control*, Special Issue on *Analysis and Control of Multi-Agent Dynamic Systems, September 2007*, pp. 584–596.
30. F. Qureshi and D. Terzopoulos, Distributed coalition formation in visual sensor networks: A virtual vision approach, *3rd IEEE International Conference on Distributed Computing in Sensor Systems*, 2007.