Dynamic Camera Assignment and Handoff

Bir Bhanu and Yiming Li

12.1 Introduction ................................................................. 338
12.2 Technical Approach ...................................................... 339
  12.2.1 Motivation and Problem Formulation .................. 339
  12.2.2 Game Theoretic Framework .............................. 339
    12.2.2.1 Computation of Utilities ....................... 340
    12.2.2.2 Bargaining among Cameras .................. 342
  12.2.3 Experimental Results ........................................ 345
    12.2.3.1 Data .................................................. 345
    12.2.3.2 Criteria for Camera Assignment and Handoff .......... 346
    12.2.3.3 Evaluation Measurements ....................... 348
    12.2.3.4 Analysis for Experiments with Different Criteria .......... 348
    12.2.3.5 Convergence of Results for Bargaining .......... 354
    12.2.3.6 Comparison with Another Related Approach .......... 355
12.3 Summary .................................................................... 358
12.4 Bibliographical and Historical Remarks ...................... 360
References .............................................................................. 360
12.1 INTRODUCTION

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 the cameras may have overlapping or nonoverlapping FOVs, seamless tracking of moving objects can be achieved by exploiting the handoff capability of multiple cameras. 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 that is being used to track it. This will provide a better situation assessment of the environment under surveillance. 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. It is evident that the manual camera handoff will become unmanageable when the number of cameras is large. Therefore, we need to develop surveillance systems that can automatically carry out the camera assignment and handoff task.

In this chapter, we provide a new perspective to the camera handoff problem that is based on game theory. The merit of our approach is that it is independent of the topology of how the cameras are placed. When multiple cameras are used for tracking and where multiple cameras can “see” the same object, the algorithm can automatically provide an optimal as well as stable solution of the camera assignment quickly. Since game theoretic approach allows dealing with multiple criteria optimization, we are able to choose the “best” camera based on multiple criteria that are selected a priori. The detailed camera calibration or 3D scene understanding is not needed in our approach.

Our approach differs from the earlier traditional approaches. We propose a game theoretic approach for camera assignment and handoff using the vehicle-target model (Arslan et al. 2007). We model camera assignment and handoff as a multiplayer game and allow for both coordination and conflicts among these players. Multiple criteria, which are used to evaluate the tracking performance, are used in the
utility functions for the objects being traced. The equilibrium of the game provides the solution of the camera assignment.

The following sections are devoted to describing the proposed game-theory-based method. Section 12.2 formulates the camera assignment and handoff problem and then constructs the utilities and bargaining steps. The implementation of this approach and some experimental results are also shown in this section. The conclusions are given in Section 12.3. Some related works in this area are reviewed in Section 12.4.

12.2 TECHNICAL APPROACH

12.2.1 Motivation and Problem Formulation

Game theory can be used for analyzing the interactions as well as conflicts among multiple agents. Analogously, in a video sensor network, collaborations as well as competitions among cameras exists simultaneously. This concept enlightens us to view the camera assignment problem in a game theoretic manner. The interactive process is called a game, while all the participants of the game are called players, who strive to maximize their utilities. 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 (Arslan et al. 2007) 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.

12.2.2 Game Theoretic Framework

Game theory involves utility, which refers to the amount of “welfare” an agent derives in a game (Myerson 1991, Li and Bhanu 2008). 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 the persons assigned to it based on the user-supplied criteria.
3. **Person utility**: How well the person is satisfied with being tracked by the current selected camera.

Our objective is to maximize the global utility as well as to make sure that the "best" camera tracks each person. 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, view, etc.) are computed. Various utilities (camera utility, person utility, predicted person utility, and global utility) are computed based on the user-supplied criteria and *bargaining process* among available cameras are executed based on the prediction of person utilities (the so-called *predicted person utility*) in each step. The results obtained from the strategy execution are in turn used for updating the camera utilities and person utilities until the strategies converge. Finally, those cameras with the highest converged probabilities will be used for tracking and 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 for our approach and their descriptions are given in Table 12.1.

### 12.2.2.1 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
the person utility \( P_i U \) is said to be aligned with the global utility \( U_g \), where \( a_{-i} \) stands for the assignments for persons other than \( P_i \), that is, \( a_{-i} = (a_1, ..., a_{-i}, a_{i+1}, ..., a_{N_p}) \). So, we can also express \( a \) as \( a = (a_i, a_{-i}) \).

We define the global utility as

\[
U_g(a) = \sum_{C_j \in C} U_{C_j}(a)
\]

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. Now, we define the person utility as

\[
U_{P_i}(a) = U_g(a_i, a_{-i}) - U_g(C_0, a_{-i}) = U_{C_j}(a_i, a_{-i}) - U_{C_j}(C_0, a_{-i})
\]
The person utility $U_P(a)$ can be viewed as a marginal contribution of to the global utility, that is, the difference of the utility gained by assigning camera to track person compared with is tracked by a virtual camera (no camera). To calculate (12.3), we have to construct a scheme to calculate the camera utility $U_C(a)$. We assume that there are $N_{Cr}$ criteria to evaluate the quality of a camera used for tracking an object. Thus, the camera utility can be built as

$$U_C(a, a_i) = \sum_{s=1}^{n_p} \sum_{l=1}^{N_{Cr}} Cr$$  \hspace{1cm} (12.4)

where $n_p$ is the number of persons that are currently assigned to camera for tracking. Plugging (12.4) into (12.3) we can obtain

$$U_P(a_i, a_{-i}) = \sum_{s=1}^{n_p} \sum_{l=1}^{N_{Cr}} Cr_{sl} - \sum_{s=1}^{n_p} \sum_{l=1}^{N_{Cr}} Cr$$  \hspace{1cm} (12.5)

where $s \neq P_i$ means that we exclude person from the 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.

12.2.2.2 Bargaining among Cameras

As stated previously, our goal is to optimize each person’s 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 $k$th step is made according to a set of probabilities

$$p_i(k) = [p_i^1(k), \ldots, p_i^l(k), \ldots, p_i^{n_c}(k)]$$
where $n_C$ is the number of cameras that can “see” the person and $\sum_{i=1}^{n_C} p_i(k) = 1$, with each $0 \leq p_i(k) \leq 1, l = 1, \ldots, n_C$. We can generalize $p_i(k)$ to be

$$p_i(k) = [p_1^i(k), \ldots, p_l^i(k), \ldots, p_{n_C}^i(k)]$$

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

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

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 $\bar{U}_P(k)$: 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 (12.5), 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 $\bar{U}_P(k)$ is calculated by (12.6)

$$
\bar{U}_P^l(k+1) = \begin{cases} 
   \bar{U}_P^l(k) + \frac{1}{p_i^l(k)} (U_P^l(a(k))) - \bar{U}_P^l(k), & a_i(k) = A_i^l \\
   \bar{U}_P^l(k), & \text{otherwise}
\end{cases} 
$$

with the initial state $\bar{U}_P^l(1)$ to be assigned arbitrarily as long as it is within the reasonable range for $\bar{U}_P^l(k)$, 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 (Arslan et al. 2007),

$$
P_i(k)^T \bar{U}_P(k) + \tau H((p_i)k)
$$

(12.7)
where

\[ H(p_i(k)) = -p_i(k)^T \log(p_i(k)) \]  \hspace{1cm} (12.8)

is the entropy function and \( \tau \) is a positive parameter belonging to \([0,1]\) that controls the extent of randomization. The larger the \( \tau \) is, the faster the bargaining process converges; the smaller the \( \tau \) is, the more accurate result we can get. So, there is a trade-off when selecting the value of \( \tau \) and we select \( \tau \) to be 0.5 in our experiments. The solution of (12.7) is proved (Arslan et al. 2007) to be

\[ p_i(k) = \frac{e^{(1/\tau)C_i^k(k)}}{e^{(1/\tau)C_i^k(k)} + \cdots + e^{(1/\tau)C_i^k(k)}} \]  \hspace{1cm} (12.9)

After several steps of calculation, the result of \( p_i \) tends to converges (refer to Figure 12.8). Thus, we finally get the stable solution, which is proved to be at least suboptimal (Arslan et al. 2007). This overall algorithm is summarized in Algorithm as follows:

\textbf{Algorithm: Game Theoretic Camera Assignment and Handoff}

\textbf{Input:} Multiple video streams.
\textbf{Output:} A probability matrix according to which camera assignments are made.

\textbf{Algorithm description:}

\begin{itemize}
  \item At a given time, perform motion detection and get the selected properties for each person that is to be tracked.
  \item For each person and each camera, decide which cameras can "see" a given person \( P_i \).
  \item For those that can “see” the person \( P_i \), initialized the predicted person utility vector \( U_{P_i}(1) \).
\end{itemize}

\textbf{Repeat}

1. Derive the \( C_{r_i} \) for each available camera.
2. Compute the camera utilities \( U_{C_i}(a) \) by (4).
3. Compute the person utilities \( U_{P_i}(a) \) by (5).
4. Compute the predicted person utilities \( U_{P_i}(k) \) by (6).
5. Derive the strategy by \( P_i(k) \) using (9).

\textbf{Until} the strategies for assignment converge.

\begin{itemize}
  \item Do the camera assignment and handoff based on the converged strategies.
\end{itemize}
12.2.3 Experimental Results

12.2.3.1 Data

In this section, we test the proposed approach for both a single person and two persons who are walking through three Axis 215 PTZ cameras. The experiments are carried out with no camera calibration needed. The cameras are placed arbitrarily. An illustration for the camera configuration for the experiment is shown in Figure 12.1. To fully test whether the proposed approach can help to select the “best” camera based on the user-supplied criteria, some of the FOVs of these cameras are allowed to intersect intentionally while some of them are nonoverlapping. This is important for tracking various people in a camera network.

A person observer selects the walking person manually when he first enters the FOV of a camera and detected by background subtraction around the edges when he leaves and re-enters the FOV (we suppose that there are no doors in the FOVs). The tracking is done by using the Continuous Adaptive Mean-Shift Algorithm proposed by Bradski.

![Figure 12.1](image.png)  
**Figure 12.1** Camera configuration in our experiments.
in Bradski (1998). Different persons are identified by calculating the correlation of the hue histograms of the pixels inside their bounding boxes using the OpenCV function CompareHist (http://opencv.willowgarage.com/wiki/CvReference). In our experiment, we compare the colors of the upper bodies (around 1/2 size of the bounding box) first, when the colors of the upper bodies are similar, then, we continue to compare the colors of the lower body and identify persons by the color combination of the upper body and the lower body.

For the sake of privacy, we perform only simulations for large numbers of cameras and persons instead of doing the experiments with real data.

12.2.3.2 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 $\lambda$ is the threshold for best observation, that is, when $r=\lambda$ this criterion reaches its peak value, where $r = \frac{\text{# of pixels inside the boundary box}}{\text{# of pixels inside the image plane}}$.

   $$\text{Crt}_{11} = \begin{cases} \lambda r, & \text{when } r < \frac{1}{\lambda} \\ \frac{1-r}{1-\lambda}, & \text{when } r \geq \frac{1}{\lambda} \end{cases}$$  

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.
Dynamic Camera Assignment and Handoff

\[ Crt_{i2} = \frac{\sqrt{(x - x_c)^2 + (y - y_c)^2}}{\frac{1}{2} \sqrt{x_c^2 + y_c^2}} \]  
(12.11)

where \((x, y)\) is the current position of the person and \((x_c, y_c)\) 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, i.e. when \(R = \xi\) the view of the person is the best, where

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

\[ Crt_{i3} = \begin{cases} \xi r, & \text{when } R < \frac{1}{\xi} \\ 1 - \frac{R}{1 - \xi}, & \text{when } R \geq \frac{1}{\xi} \end{cases} \]  
(12.12)

4. Combination of criteria 1, 2, and 3, which is called the combined criterion is given by the following equation,

\[ Crt_{i4} = \sum_{m=1}^{3} w_m Crt_{im} \]  
(12.13)

where \(w_m\) are the weights for different criteria.

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

In our experiments, we give value to the parameters empirically. \(\lambda = (1/15), \xi = (1/6), w_1 = 0.2, w_2 = 0.1, \) and \(w_3 = 0.7\). This is because we prefer a frontal view of the person whenever it is available, while a continuous tracking is the bottom line when the frontal view is not detected.
12.2.3.3 Evaluation Measurements

In our experiments, the bottom line is to track walking persons seamlessly whenever they appear in the FOV of any of the cameras. In the case where more than one camera can “see” the persons, those ones that can “see” the persons’ face are always the most preferable. Based on this goal, if we define the error in our experiments as either failing to track a person or failing to get the frontal-view of the person whenever it is available. The performances for using criterion 1, criterion 2, or criterion 3 alone in a two-person experiment are 25.56%, 10.00%, and 30.00% respectively. We can notice that based on our error definition, single criterion incurs a high error rate. Hence, we introduce the combined criterion to overcome this error rate.

12.2.3.4 Analysis for Experiments with Different Criteria

• A single-person case

Figure 12.2 gives the camera handoffs based on the combined criterion in a single-person experiment. The camera with a dark bounding box is the one to be chosen. As shown in Figure 12.2, a frontal-view person (whenever it is available) is selected for most of the frames. Sometimes, the frontal view is not selected because the face is not detected. In some frames, such as in Figure 12.2b, although the frontal view is available, the person is too close to the edge of the image or the size of the person is far from the “good size” threshold. In this case, the system will choose some other available camera. All the handoffs and interesting events are listed in Table 12.2, where we use E denoting that the person is entering the FOV of a camera while L denoting that the person is leaving the FOV of a camera. A implies that the camera can see the object and, thus, it is available for tracking, while N implies that there is no object in the FOV of a camera. The last column in Table 12.2, Used, gives the camera that is selected to track the person. There are altogether 600 frames. It shows that camera handoff is carried out correctly especially when the person is entering or leaving the FOV of some cameras.
Figure 12.2 (See color insert following page xxx.) Camera assignments and handoffs in the 1 person 3 cameras case. The camera in which the bounding box is the dark color is selected to track the person.
<table>
<thead>
<tr>
<th>Frame</th>
<th>Camera 1</th>
<th>Camera 2</th>
<th>Camera 3</th>
<th>Used</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>A</td>
<td>A</td>
<td>N</td>
<td>1</td>
</tr>
<tr>
<td>69</td>
<td>A</td>
<td>A</td>
<td>E</td>
<td>1</td>
</tr>
<tr>
<td>88</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>2</td>
</tr>
<tr>
<td>120</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>3</td>
</tr>
<tr>
<td>121</td>
<td>L</td>
<td>A</td>
<td>A</td>
<td>3</td>
</tr>
<tr>
<td>131</td>
<td>E</td>
<td>A</td>
<td>A</td>
<td>3</td>
</tr>
<tr>
<td>209</td>
<td>A</td>
<td>L</td>
<td>A</td>
<td>3</td>
</tr>
<tr>
<td>226</td>
<td>A</td>
<td>N</td>
<td>L</td>
<td>1</td>
</tr>
<tr>
<td>357</td>
<td>A</td>
<td>N</td>
<td>E</td>
<td>3</td>
</tr>
<tr>
<td>379</td>
<td>A</td>
<td>E</td>
<td>A</td>
<td>3</td>
</tr>
<tr>
<td>409</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>1</td>
</tr>
<tr>
<td>427</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>2</td>
</tr>
<tr>
<td>483</td>
<td>L</td>
<td>A</td>
<td>A</td>
<td>2</td>
</tr>
<tr>
<td>556</td>
<td>E</td>
<td>A</td>
<td>A</td>
<td>2</td>
</tr>
<tr>
<td>585</td>
<td>A</td>
<td>A</td>
<td>L</td>
<td>2</td>
</tr>
<tr>
<td>600</td>
<td>A</td>
<td>A</td>
<td>N</td>
<td>END</td>
</tr>
</tbody>
</table>

Note: **A**, available; **E**, entering; **L**, leaving; **N**, not available.

A more detailed discussion for choosing different criteria is analyzed in the two person case as discussed in the following text.

- A two-person case

Different experiments are carried out to compare the results for using the three different single criteria mentioned previously with the combined criterion. The weights we use to combine the three criteria, in our experiments, are 0.2, 0.1, and 0.7 respectively, as stated previously. To make it convenient for a comparison, we show the tracking results of other cameras as well, no matter whether they are selected for tracking or not. Similar to the single person experiment, the cameras, for which the bounding boxes are drawn in dark color, are selected for tracking while the ones in the light color are not as good as the dark ones.

A comparison for using criterion 1, criterion 2, and criterion 3 respectively at two time instants is shown in Figure 12.3. We can
Figure 12.3 (See color insert following page xxx.) A comparison for using different criteria. The left column and the right column are for two time instants respectively. The first row through the third row are using criterion 1 to criterion 3, respectively.

observe that in Figure 12.3a through c uses criterion 1 through 3 in time instant 1 while Figure 12.3d through f uses criterion 1 through 3 at time instant 2. It can be noticed from Figure 12.3d that the problem for using criterion 1 only is that when the objects are getting
close to the cameras, the size of the bounding box will increase to a non-desirable size for observation any more. Meanwhile, there are often some cases that when the person is entering the scene, its size is not small but only part of the body is shown, which should not be preferred as well if some other cameras can give a full view of the body. Thus, we introduce criterion 2, considering the relative position of the objects in the FOVs of the cameras. The closer the centroid of the person is to the center of the FOV, the higher the camera utility is generated. We can observe that when applying criterion 2 in Figure 12.3e, the camera with the object near the center is chosen and we can, thus, obtain a higher resolution of the person compared with the results for using criterion 1 in Figure 12.3d. However, the problem for using criterion 1 or criterion 2 only is that in many frames we reject the cameras that can see a person’s face, which is of general interest. This case is shown in Figure 12.3a, b, and d. To solve this problem, we come up with criterion 3. So, when applying criterion 3, we can obtain a more desirable camera with a frontal view of the person in Figure 12.3c and f. Whereas criterion 3 can successfully select a camera with a frontal-view person, it may fail to track a person when no face can be detected. For instance, as shown in Figure 12.3f, although the person is in the FOV of some camera, it is “lost” based on criterion 3.

So, finally, we come up with a weighted combination of these three criteria. As stated previously, we use 0.2, 0.1, and 0.7 as the weights for these three criteria respectively so that, in most cases, the system will choose the camera that can “see” a person’s face. For those frames where there is a person with no face detected, the combination criterion can also provide a “best” camera based on criteria 1 and 2 and, thus, realize a continuous tracking. All the camera handoffs, when applying the combined criterion, are shown in Figure 12.4a through i. The error rate (as defined in Section 12.2.3.3) in this case is 5.56%. This combined criterion provides camera assignments and handoffs with a minimum error rate among the four criteria defined in Section 12.2.3.2. Camera utilities, person utilities, and the corresponding assignment probabilities for the using the combined criterion is shown in Figure 12.5, where Probability[i][j] stands for the probability that Camera j is assigned to track Person i.
In an n-camera n-person case, we can expect that the number of iteration of the proposed approach will go up much slower than that of the exhaustive approach. So, the computational time-saving advantage of the proposed approach will be more obvious as the situation of the task becomes more complicated, as shown in our simulation result in Figure 12.6.

Figure 12.4 (See color insert following page xxx.) All camera handoffs when applying the combined criterion for 2 persons and 3 cameras case.
In our experiments, in most cases, the probabilities for making the assignment profile converges (with $\varepsilon < 0.05$, where $\varepsilon$ is the difference between the two successive results) within 5 iteration. So, we use 5

Figure 12.5 (See color insert following page xxx.) Utilities and assignment probabilities for each processed frame when using the combined criterion. $\text{Probability}[i][j]$ stands for the probability that Camera $j$ is selected to track Person $i$.

12.2.3.5 Convergence of Results for Bargaining

In our experiments, in most cases, the probabilities for making the assignment profile converges (with $\varepsilon < 0.05$, where $\varepsilon$ is the difference between the two successive results) within 5 iteration. So, we use 5
as the iteration threshold when bargaining. Thus, for those cases that will not converge within 5 iterations, there may be an assignment error based on the unconverged probabilities, as discussed in Section 12.2.3.2. In Figure 12.7, we plot the number of iteration with respect to every processed frame. It turns out that the average iteration number for the case in our experiment is 1.37. As the numbers of persons and cameras increase, this bargaining system will save a lot of computational cost to get the optimal camera assignments. A typical convergence for one of the assignment probabilities in a bargaining among cameras is given in Figure 12.8.

**12.2.3.6 Comparison with Another Related Approach**

In this section, we will compare our approach with another camera assignment approach by Jo and Han in (2006). The authors perform
camera handoffs by calculating the COR (the co-occurrence to occurrence ratio). We will call this the COR approach.

In Jo and Han (2006), the mean probability that a moving object is detected at a location $x$ in the FOV of a camera is called an occurrence at $x$. The mean probability that moving objects are simultaneously detected at $x$ in the FOV of one camera and $x'$ in the FOV of another camera is called a co-occurrence of $x$ and $x'$. The COR approach decides whether two points are in correspondence with each other by calculating the 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 handoff to another camera that has its corresponding point. However, the COR approach in Jo and Han (2006) 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.

Experiments have been done to compare the COR approach with our approach for the 1 person 3 cameras case and the 2 persons 3 cameras case.
• The 1 person 3 cameras case

The handoff process by using the COR approach and the corresponding frames by using our approach (may not be the handoff frames) are shown in Figure 12.9. In Figure 12.9f through h, the COR approach

Figure 12.9 (See color insert following page xxx.) Two camera handoffs by using the co-occurrence to occurrence ratio (COR) approach and the comparison with our approach. The left column are the results by our approach and the right column are the results by the COR approach.
switches to camera 1, while our proposed approach sticks to camera 2 to get the frontal view of the person. The COR approach needs a time period to construct the correspondence between views of different. As stated earlier, we let this period to be 10 frames. As a result, there is some time delay for the handoff. For instance, in Figure 12.9a through d, our approach has already switched to camera 3 in Figure 12.9a as long as the size of the person is unacceptable, but the COR approach does this in Figure 12.9d.

- The 2 person 3 cameras case

In Figure 12.10, we show the comparison of results in this case. We can notice that the COR approach produces more errors. Whenever there is occlusion in one of the FOVs, there is a high probability that the COR approach will provide the wrong points in another FOV and this causes loss of track for some persons. Examples are Figure 12.10b, d, f. By the comparison, we can notice that the COR approach can only switch the camera to another one when the person is about to leave the FOV, but cannot select the “best” camera based on other criteria. So, the number of handoffs by our approach is larger than that of the COR approach (see Table 12.3). If we use the definition of error as stated in the Section 12.2.3.3, the error rates for these two cases are compared in Table 12.3.

### 12.3 SUMMARY

In this paper, we proposed a new principled approach based on game theory for camera assignment and handoff problem. The approach is independent of the spatial and geometrical relationships among the cameras. It is robust with respect to multiple criteria for tracking are considered. The key merit of the proposed approach is that we use the game theory framework with bargaining mechanism for camera assignment in a video network so that we can obtain a stable solution with a small number of iterations. This makes our approach computationally more efficient and robust with respect to other existing approaches,
such as Jo and Han (2006). Future work will allow communications among cameras, which will make the computational framework and computational resources decentralized and distributed.
There have been many papers discussing approaches for camera assignments in a video network. Javed et al. (2000) focus on finding out the limits of overlapping FOVs of multiple cameras. Park et al. (2006) create distributed look-up tables according to how well the cameras can image a specific location. Jo and Han (2006) construct a handoff function by computing the ratio of co-occurrence to occurrence for several pairs of points in the FOVs of two corresponding cameras. This kind of approaches rely on obtaining the spatial topology of the camera network and calculating the geometrical relationships among cameras, which tends to be quite complicated when the topology becomes complex and it is difficult to learn the topology based on the random traffic patterns. Statistics-based methods (Kettnaker and Zabih 1999, Chang and Gong 2001, Kang et al. 2003, Javed et al. 2005, Song and Roy-Chowdhury 2007) provide an optimal solution with respect to object trajectories, while other factors, such as orientation, shape, and face, which are also very important for tracking, are not considered. Also, many researchers have used calibrated cameras, an example is Cai and Aggarwal (1999).

**REFERENCES**


http://opencv.willowgarage.com/wiki/CvReference#Histograms


AUTHOR QUERY

[AQ1] Please provide the significance of the usage of bold in Table 12.2.

[AQ2] Please check whether the inserted source line is correct.

[AQ3] Please provide the initials for the author Khan.