UTILITY-BASED DYNAMIC CAMERA ASSIGNMENT AND HAND-OFF IN A VIDEO NETWORK

Yiming Li and Bir Bhanu

Center for Research in Intelligent Systems
University of California at Riverside
Riverside, CA 92521, USA
E-mail: yimli@vislab.ucr.edu; bhanu@cris.ucr.edu

ABSTRACT

In this paper we propose an approach for multi-camera multi-person seamless tracking that allows camera assignment and hand-off based on a set of user-supplied criteria. The approach is based on the application of game theory to camera assignment problem. Bargaining mechanisms are 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. Experiments for multi-camera multi-person tracking are provided. Several different criteria and their combination of them are carried out and compared with each other to corroborate the proposed approach.

***Index Terms*—**Camera Network, Camera Selection, Detection, Multi-Camera Surveillance, Tracking

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

In this paper, we provide a new perspective to the camera hand-off 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.

The following sections are devoted to describing the proposed game-theory-based method. Section 2 describes the related work and contributions of this work. Section 3 formulates the camera assignment and hand-off problem and then constructs the utilities and bargaining steps. Section 4 discusses the implementation of this approach and shows the experimental results. The final conclusions are given in Section 5.

2. RELATED WORK AND OUR CONTRIBUTIONS

There have been many papers discussing approaches for camera assignments in a video network. Javed *et al.* [1] focus on finding out the limits of overlapping FOVs of multiple cameras. Park *et al.* [2] create distributed look-up tables according to how well the cameras can image a specific location. Jo and Han [3] construct a hand-off function by computing the ratio of co-occurrence to occurrence for all pairs of points in two successive views. This kind of approach relies 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 it based on the random traffic patterns. Statistics-based methods [4, 5, 6, 7, 8] provide an optimal solution with respect to object trajectories, while other factors, such as orientation, shape, face and etc., which are also very import for tracking, are not considered. Also, many researches have used calibrated cameras, and an example is [9].

Our approach differs from the above traditional approaches. We propose a game theoretic approach for camera assignment and hand-off using the vehicle-target model [10]. We model camera assignment and hand-off as a multi-player 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 tracked. The

978-1-4244-2665-2/08/$25.00 ©2008 IEEE

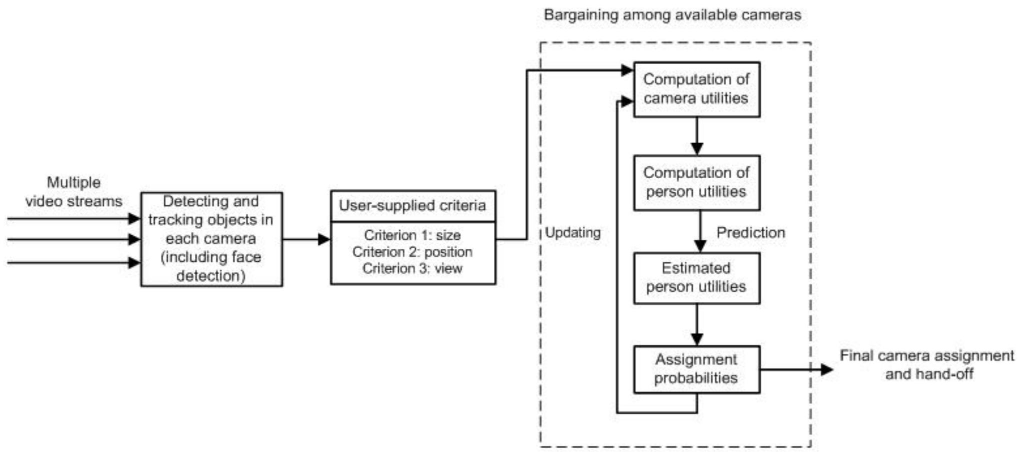

Figure 1: Game theoretic framework for camera assignment and hand-off.

equilibrium of the game provides the solution of the camera assignment.

3. TECHNICAL APPROACH

3.1. 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, communications as well as competitions among cameras exist simultaneously. This 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 multi-player game, with the available cameras being the players. If there are multiple persons in the system, this becomes a multiple of multi-player game being played simultaneously.

Vehicle-target assignment [10] is a multi-player 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.

3.2. Game Theoretic Framework

Game theory involves *utility*, which refers to the amount of 'welfare' an agent derives in a game [11]. We are concerned with three different utilities: *global utility*, the overall degree of satisfaction for tracking performance, *camera utility*, how well a camera is tracking the persons assigned to it based on the user supplied criteria, and *person utility*, how well the 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.

An overview of the approach is illustrated in Figure 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, view, etc.) are computed. Various utilities (camera utility, 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 in 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 will be used for tracking and this assignment of persons to the "best" cameras leads to the solution of the hand-off problem in multiple video streams.

A set of symbols are used in the discussion for our approach and their descriptions are given in Table 1.

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 as 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.

Table 1: *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, $A_i = \{a_1, a_2, ..., a_{n_C}\}$
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_{-i}	The assignment of cameras for the persons excluding person i
a	Assignment of cameras for all persons, $a = (a_i, a_{-i})$
$U_{cj}(a)$	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, $\bar{U}_{P_i}(k) = [\bar{U}^1_{P_i}(k), ..., \bar{U}^l_{P_i}, ..., \bar{U}^{n_C}_{P_i}(k)]^T$, where $\bar{U}^l_{P_i}$ is the predicted person utility for P_i if camera a_l is used
$p_i(k)$	Probability of person i's assignment at step k, $p_i(k) = [p^1_i(k), ..., p^l_i(k), ..., p^{n_C}_i(k)]$, where $p^l_i(k)$ is the probability for camera a_l to track person P_i

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

$$U_{P_i}(a'_i, a_{-i}) < U_{P_i}(a''_i, a_{-i}) \Leftrightarrow U_g(a'_i, a_{-i}) < U_g(a''_i, a_{-i}) \quad (1)$$

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., $a_{-i} = (a_1, ..., a_{i-1}, a_{i+1}, ..., a_{N_P})$. We define the global utility as

$$U_g(a) = \sum_{C_j \in C} U_{C_j}(a) \quad (2)$$

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:

$$\begin{aligned} U_{P_i}(a) &= U_g(a_i, a_{-i}) - U_g(C_0, a_{-i}) \\ &= U_{C_j}(a_i, a_{-i}) - U_{C_j}(C_0, a_{-i}) \end{aligned} \quad (3)$$

The person utility $U_{p_i}(a)$ can be viewed as a marginal contribution of P_i to the global utility. To calculate (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

$$U_{C_j}(a_i, a_{-i}) = \sum_{s=1}^{n_P} \sum_{l=1}^{N_{Crt}} Crt_{sl} \quad (4)$$

where n_P is the number of persons that are currently assigned to camera C_j for tracking. Plugging (4) into (3) we can obtain

$$U_{P_i}(a) = \sum_{s=1}^{n_P} \sum_{l=1}^{N_{Crt}} Crt_{sl} - \sum_{\substack{s=1 \\ s \neq P_i}}^{n_P} \sum_{l=1}^{N_{Crt}} Crt_{sl} \quad (5)$$

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

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 k^{th} step is made according to a set of probabilities

$$p_i(k) = [p^1_i(k), ..., p^l_i(k), ..., p^{n_C}_i(k)]$$

where n_C is the number of cameras that can "see" the person P_i. At each bargaining step, we will assign a person to the camera which 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 (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 (6):

$$\bar{U}^l_{P_i}(k+1) = \begin{cases} \bar{U}^l_{P_i}(k) + \dfrac{1}{P^l_i(k)}(U_{P_i}(a(k)) - \bar{U}^l_{P_i}(k)), & a_i(k) = A^l_i \\ \bar{U}^l_{P_i}(k), & \textit{otherwise} \end{cases} \quad (6)$$

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 $l=1,..,n_c$. Once these predicted person utilities are calculated, it can be proved that the equilibrium for the strategies lies in the probability distribution [10],

$$p_i^l(k) = \frac{e^{\frac{1}{\tau}\bar{U}_{P_i}^l(k)}}{e^{\frac{1}{\tau}\bar{U}_{P_i}^1(k)} + \ldots + e^{\frac{1}{\tau}\bar{U}_{P_i}^{n_C}(k)}} \tag{7}$$

where τ is a small number between 0 and 1. After several steps of calculation, the result of $p_i(k)$ tends to converges. Thus, we finally get the stable solution, which is proved to be at least suboptimal.

4. EXPERIMENTAL RESULTS

In this section, we describe experiments to evaluate the performance to corroborate the proposed approach.

4.1. Data

In our experiments, we test the proposed approach for both a single person and two persons who are walking through three Axis 215 PTZ cameras. The cameras are placed arbitrarily. 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 non-overlapping. This is important for tracking various people in a camera network.

The video is 20 seconds long for tracking a single person. For two persons tracking the video is 15 seconds long and the frame rate is 30 f/s, so we have 450 frames for every camera. To speed up the processing, we apply the bargaining mechanism every 5 frames, i.e. 6 frames are processed per second.

4.2. Criteria for Camera Assignment and Hand-off

A number of criteria, including human biometrics, can be used for camera assignment and hand-off. 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, as measured by the number of pixels on the person.
(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 a video frame.
(3) The view of the person, as measured by the ratio of the number of pixels on the detected face to that of the maximum bounding box.
(4) Combination of criterion (1), (2) and (3), which is called the combined criterion.

All these criteria are normalized for calculating the corresponding camera utilities.

4.3. Evaluation Measurements

In our experiments, the bottom line is to track the 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 objects, the one that can "see" the person's face is always the most preferable. Given that the tracking is ideal, when single criterion is used, the average error rates for using criterion 1 (number of pixels on the person), criterion 2 (the relative position of the objects in the FOVs of the cameras) and criterion 3 (the view of the person) are 2.22%, 1.11% and 2.22% respectively. However, based on our goal to do the camera assignments as discussed previously in this subsection, we can re-define the error in our experiments as either failing to track a person or failing to get the frontal-view person whenever it is available. The performance for all these cases in a two-person experiment is given in Table 2.

Table 2: *Performance for using different criteria*

CRITERION USED	ERROR RATE
Criterion 1	25.56%
Criterion 2	10.00%
Criterion 3	30.00%
Combined criterion	5.56%

4.4. Analysis of Experiments with Different Criteria

4.4.1. A Single-Person Case

Figure 2 shows some typical camera hand-offs based on the combined criterion in a single-person experiment. The camera with a yellow bounding box is the one to be chosen. As shown in the figure, a frontal-view person (whenever it is available) is selected for most of the frames. All the hand-offs are listed in Table 3, where we use **E** to denote that the person is entering the FOV of a camera, while **L** denotes that the person is leaving the FOV of a camera. **A** means that the camera can see the object and, thus, it is available for tracking, while **N** stands for that there is no object in the FOV of a camera.

The video starts at 07:00:24.42 and ends at 07:00:44:42. We record all the interesting events in Table 3. It shows that camera hand-off is carried out correctly even when the person is entering or leaving the FOV of some cameras.

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

4.4.2. A Two-person Case

Table 4 gives a general description of the videos in each of the three cameras and the number of hand-offs based on the combined criterion.

Table 3: *Hand-offs among 3 cameras during 20 seconds (A: available; E: entering; L: leaving; N: not available)*

Time	Camera0	Camera1	Camera2	Used
07:00:24.42	A	A	N	0
07:00:24.81	A	A	N	1
07:00:26.47	A	A	E	1
07:00:29:00	A	A	A	0
07:00:31.39	A	L	A	0
07:00:32.97	A	N	L	0
07:00:35.00	A	N	E	0
07:00:35.38	A	N	A	2
07:00:37.33	A	E	A	2
07:00:37.43	A	A	A	0
07:00:40.43	A	A	A	1
07:00:40.67	L	A	A	1
07:00:43.11	E	A	A	1
07:00:44.20	A	A	L	1
07:00:44.42	N	N	N	END

Different experiments are carried out for using the three different single criterion mentioned previously and a combined criterion. The weights we use to combine the three criteria in our experiments are 0.2, 0.1 and 0.7 respectively, since we are always expecting to see a person's face whenever it is possible. To make it convenient for a comparison, we show the tracking results of other cameras as well, no mater whether they are selected for tracking or not. The camera for which the bounding box(es) is (are) drawn in blue is (are) selected to be used for tracking while the ones with red or green bounding boxes are decided to be not as good as the blue one.

A comparison for using criterion 1, criterion 2 and criterion 3 respectively at two time instants is shown in Figure 3. Figure 3(a) to 3(c) are using criterion 1 to 3 at time instant 1 while (d) to (f) are using criterion 1 to 3 at time instant 2. It can be noticed from Figure 3(d) 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, while the resolution is not that high, making the objects not clear enough for recognition. Meanwhile, there are often some cases when a person is entering the scene, its size is not small but only part of the body is shown, which should not be preferred if some other cameras can give a full view of the body. Thus, we introduced 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 this when applying criterion 2 in Figure 3(e), the camera with the object near to the center is chosen and we can, thus, obtain a higher resolution on the person compared with the results obtained from using the criterion 1 in 3(d). However, the problem with using criterion 1 or criterion 2 only, is that in many frames, we reject the camera(s) which can see a person's face, which is of general interest. This case is shown in Figure 3(a) (b) and (d) To solve this problem, we come up with criterion 3 (the view of the person). So, when applying criterion 3, we obtain a more desirable camera with a frontal view of the person in Figure 3(c) 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. As shown in Figure 3(f), 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 and the system will choose the camera which can "see" a person's face. For those frames where there is person without a detected face, the combination criterion can also provide a "best" camera based on criteria 1 and 2 and thus achieving continuous tracks. All the camera hand-offs when applying the combined criterion are shown in Figure 4(a) to (i). The error rate (as defined in section 4.3) in this case is 5.56%. As shown in Table 2 previously, this combined criterion provides camera assignments and hand-offs with a minimum error rate among the four criterion defined in section 4.2. Camera utilities, person utilities and the corresponding assignment probabilities for the using the combined criterion are shown in Figure 5.

Table 4: *overview of videos in each camera and hand-offs taken place.*

	No. of frames with 0 person	No. of frames with 1 person	No. of frames with 2 person	No. of frames with occlusion	No. of hand-offs based on the combined criterion
Cam0	56	22	12	0	2
Cam1	14	46	18	11	9
Cam2	44	23	17	6	6

Figure 2: Three typical camera assignments and hand-offs in a single-person case.

Figure 3: 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.

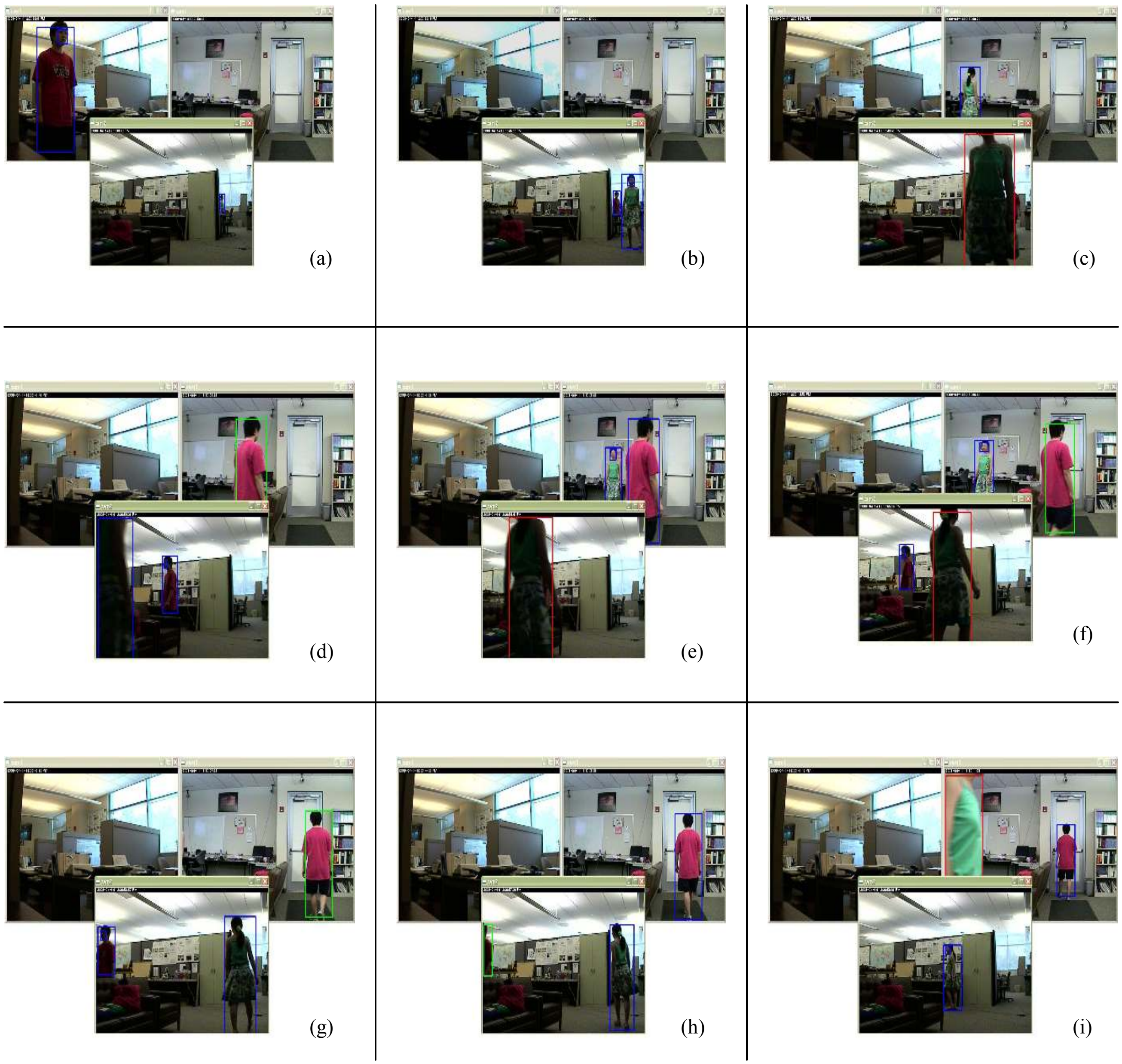

Figure 4: All camera hand-offs when applying the combined criterion.

4.5. Convergence of Results for Bargaining

In our experiments, the probabilities for making the assignment profile will converge (with $\varepsilon < 0.05$, where ε is the difference between the two successive results) within 5 iterations in most cases. So we use 5 as the iteration threshold when bargaining. In Figure 6 we plot the number of iteration with respect to every processed frame. It turns out that the average iteration number 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 the process of bargaining among cameras is given in Figure 7.

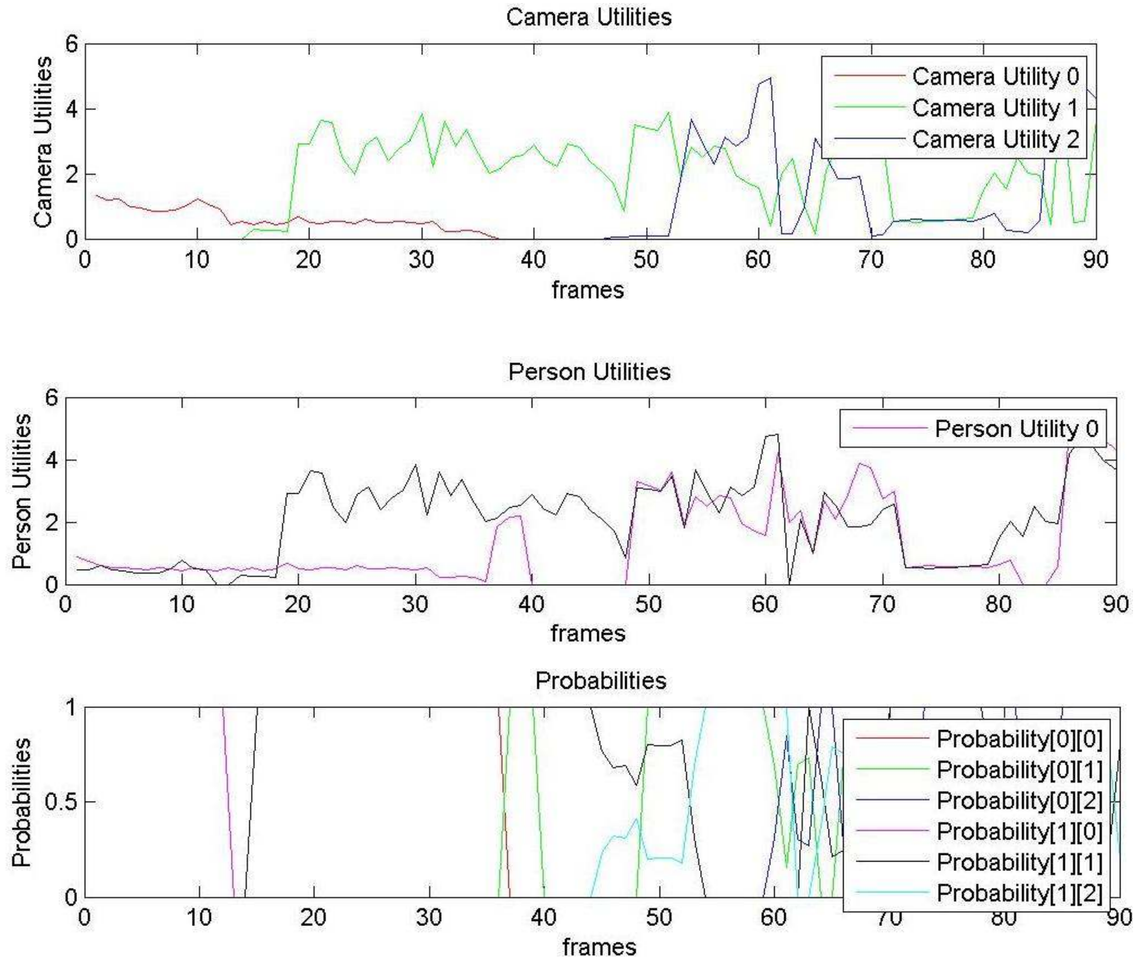

Figure 5: Utilities and assignment probabilities for each processed frame when using the combined criterion.

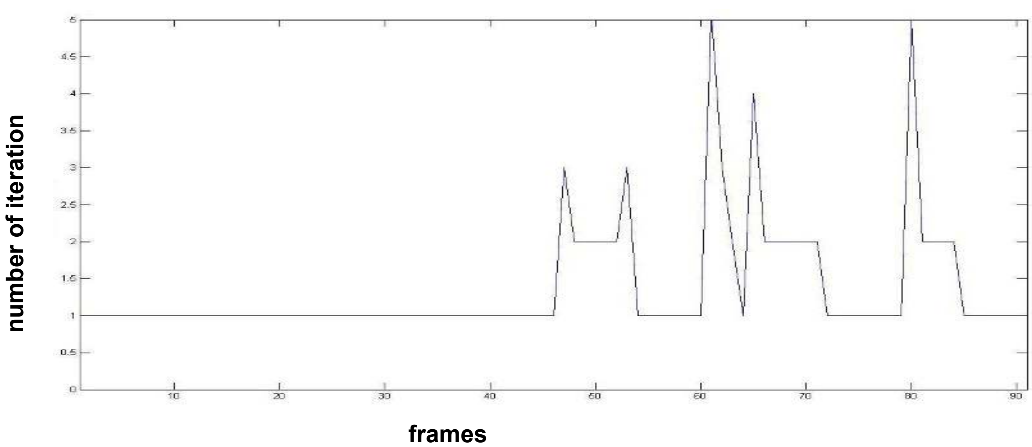

Figure 6: Number of iteration for the bargaining mechanism in each frame.

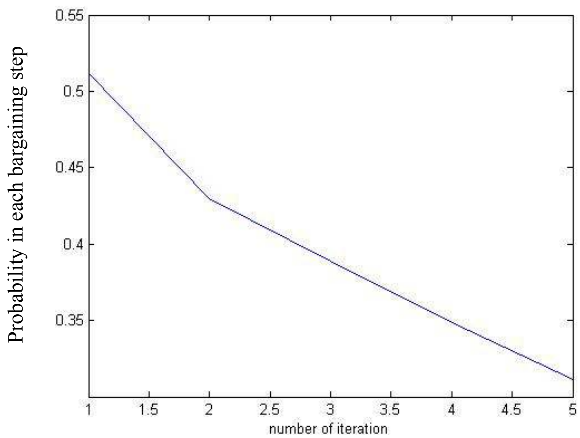

Figure 7: A Typical convergence plot obtained during the bargaining mechanism.

5. CONCLUSIONS

In this paper, we proposed a new principled approach based on game theory for camera assignment and hand-off problem. The approach is independent of the spatial and geometrical relationships among the cameras. It is robust with respect to multiple criteria for tracking that can be considered. Experiments have been carried out to test the proposed approach. Future work will extend the experiments to multi-person tracking in an environment of a large number of cameras.

6. ACKNOWLEDGMENT

This work was partially supported by NSF grants 0551741 and 0622176, and the ONR grants (DoD Instrumentation and the Aware Building). The contents of the information do not reflect the position or policy of the US Government.

7. REFERENCES

[1] O. Javed, S. Khan, Z. Rasheed and M. Shah, “Camera Hand-off: Tracking in Multiple Uncalibrated Stationary Cameras.” *IEEE Workshop on Human Motion*, pp. 113-118, 2000.

[2] J. Park, P. C. Bhat, and A. C. Kak, “A Look-up Table Based Approach for Solving the Camera Selection Problem in Large Camera Networks,” *IEEE Workshop on DSC*, 2006.

[3] Y. Jo and J. Han, “A new Approach to Camera Hand-off without Camera Calibration for the General Scene with Non-planar ground,” *ACM International Workshop on Video Surveillance and Sensor Networks,* pp. 195-202, 2006.

[4] V. Kettnaker and R. Zabih , “Bayesian Multi-camera Surveillance,” *IEEE CVPR* 1999, pp. 253-259.

[5] T. Chang and S. Gong, “Bayesian Modality Fusion for Tracking Multiple People with a Multi-Camera System,” *European Workshop on Advanced Video-based Surveillance Systems*, 2001.

[6] J. Kang, I. Cohen, and G. Medioni, “Continuous tracking within and across camera streams,” *IEEE CVPR*, Vol. 1. pp. 267-272, 2003.

[7] O. Javed, K. Shafique, and M. Shah, “Appearance modeling for tracking in multiple non-overlapping cameras,” *IEEE CVPR*, Vol.2. pp. 26-33, 2005.

[8] B. Song, A. Roy-Chowdhury, “Stochastic Adaptive Tracking In A Camera Network,” *IEEE ICCV*, pp. 1-8, 2007.

[9] Q. Cai and J. K. Aggarwal, “Tracking Human Motion in Structured Environments Using a Distributed Camera System,” *IEEE PAMI*, pp. 1241 – 1247, Nov. 1999.

[10] G. Arslan, J. R. Marden and J. S. Shamma, “Autonomous Vehicle-Target Assignment: A Game-Theoretical Formulation,” *Trans ASME*, pp. 584-596, Sept.2007.