TASK-ORIENTED CAMERA ASSIGNMENT IN A VIDEO NETWORK

Yiming Li and Bir Bhanu
Center for Research in Intelligent Systems
University of California at Riverside, Riverside, CA 92521, USA
{yimli@ee.ucr.edu, bhanu@cris.ucr.edu}

ABSTRACT

Camera assignment and hand-off are some of the key image processing problems in a video network. In this paper, we propose a new approach for camera assignment and hand-off in a video network. The camera assignment problem is modeled as a *weakly acyclic game* which allows the design of utility functions based on different user-supplied criteria. A theoretical and experimental comparison of the proposed approach with the two recently proposed approaches based on *potential game theory* and *constraint satisfaction problem* is provided. This comparison shows that the proposed approach is theoretically more general and computationally more efficient than the other approaches.

***Index Terms*—** Weakly acyclic game, Potential game, Constraint satisfaction problem, Camera assignment, Camera hand-off

1. INTRODUCTION

Due to the broad coverage of an environment and the possibility of coordination among cameras, video sensor networks have attracted much interest. Although the field-of-view (FOV) of a single camera is limited, 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.

Traditional methods in this area fall into two categories: the topology-based and the statistics-based. The former one depends on obtaining the spatial topology of the camera network and calculating the geometrical relationships among cameras [1, 2, 3], while the latter one depends on obtaining the statistical models of moving objects [4, 5]. These methods tend to be quite complicated when the topology of the camera network or the trajectory of moving objects becomes complex and it is difficult to learn based on the random traffic patterns [6]. Most of the approaches generally 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.

Recently, a few task-oriented approaches have been published [7, 8] which perform camera assignment based on the user-supplied criteria. The approach by Li and Bhanu [7] is based on the potential game theory while the approach by Qureshi and Terzopoulos [8] is based on constraint satisfaction. In this paper, we propose a new game theoretic approach, called a *weakly acyclic game*, for camera assignment. Since *weakly acyclic game* contains a larger scope of games than the potential game [7], this paper provides a more general model for camera assignment. Further, the learning process proposed in this paper makes the convergence for getting a stable solution faster than solving the constraint satisfaction problem [8]. We provide a detailed comparison of these approaches.

The rest of this paper is organized as follows: Section 2 describes the weakly acyclic game approach and the corresponding learning algorithm in detail. Section 3 provides a brief description of the potential game approach and the constraint satisfaction problem approach. Section 4 presents experimental results for the three approaches and compares them. Finally, Section 5 provides conclusions and avenues for possible future research.

2. CAMERA ASSIGNMENT AS A WEAKLY ACYCLIC GAME (WAG)

A game, in game theory, is an interactive process, while the agents in a game are called players. The welfare that a player can get from a game is called utility. The performance of the system is implied by the global utility. In our problem, if at some time instant, there are n_C cameras that can see person P_i, we say that these n_C cameras compete to track this person P_i. Thus, if we let the cameras that can see P_i be the players, then, the case that these cameras compete to track P_i can be considered as a multi-player game. For each camera, the possible actions are use (select the camera to track a person) and standby (the camera is not selected to track a person). Mathematically, let A_j be the set of actions for camera j, then $A_j = \{Use, Standby\}$. At each time instant, the actual action of camera j, a_j may equal to any of the elements in set A_j. A *better reply path* is a sequence of camera action profiles at time *t:* $a_1(t), a_2(t), ..., a_{N_C}(t)$ such that for each $a_j(t)$ there is only one action of the camera, $a_j(t) \neq a_j(t-1)$, such that the global utility $U_g(a(t)) > U_g(a(t-1))$, which implies that the system gains more from replacing

$a_j(t-1)$ for $a_j(t)$, i.e. the system has a better result by performing action . N_C is the number of cameras in the system. Since we can change the camera actions in each iteration of the learning process we can always find out a better reply path for the camera action assignment, and thus, it can be modeled as a weakly acyclic game. Since the game is finite and the path cannot cycle back on itself according to the definition, the last reply path has to be the maximal better reply path and, thus, it can be a Nash equilibrium. When a system gets to its final better reply path, the global utility will reach its maximum value as well.

To find the final better reply path, we first randomly choose a baseline action for each camera. Then, in each learning iteration, we update the camera actions with some probability ε. If it results in a better reply path, then we replace the baseline actions with this better reply path, otherwise, we keep the baseline actions to be the same as in the previous step. The probability ε can be chosen as any real number belonging to [0, 1]. The trade-off for choosing ε is that when a large ε is used, the learning process will converge quickly while it leaves the risk of losing the optimal result; on the contrary, if an ε is small enough, then we can guarantee that this learning process will access the optimal Nash equilibrium with arbitrarily high probability.

Learning Algorithm

At a given time, perform motion detection and get the relative properties of each person that is tracked.

1. For each person and each camera, decide which cameras can "see" a given person P_i.
2. For the camera which can "see" the person P_i, initialize the action of camera *j*, $a_j(0)$, randomly, and calculate the payoff $U_{C_j}(a)$. $a_j(0)$ can take value from any element of the set $A_j = \{Use, Standby\}$. Use this action as the camera's baseline action $a_j^b(0)$. Initialize the global baseline utility as $U_g^b(0) = U_g(0)$.
3. At each iteration *k*, each camera updates its action $a_j(k)$, i.e. choose another action from A_j randomly with probability ε, or stay at its baseline action $a_j^b(k-1)$ with probability $(1-\varepsilon)$, i.e.:
   - $a_j(k)$ is chosen randomly from $A_j \backslash a_j(k-1)$ with probability ε.
   - $a_j(k) = a_j^b(k-1)$ with probability $(1-\varepsilon)$.
4. If there's a better reply path, update both the baseline camera action and baseline global utility, i.e.
   - If $U_g(k) > U_g(k-1) + \delta$
     $$a_j^b(k+1) = a_j(k)$$
     $$U_g^b(k+1) = U_g(k)$$
   - If $U_g(k) \le U_g(k-1) + \delta$
     $$a_j^b(k+1) = a_j^b(k)$$
     $$U_g^b(k+1) = U_g^b(k)$$

   Where $$U_{C_j} = \sum_{i=1}^{n_P} \sum_{m=1}^{3} w_m S_{im} \tag{1}$$
   and $$U_g = \sum_{i=1}^{n_P} \sum_{m=1}^{3} L_{ijs} w_m S_{im} \tag{2}$$
   $$L_{ijs} = \begin{cases} 1, if\ C_j\ is\ selected\ for\ P_i \\ 0, \qquad otherwise \end{cases}.$$
   S_{im} are the criteria that supplied by the user, which will be designed in the experimental part and w_m *are the* weights for these criteria. n_P is the number of persons that are currently assigned to camera C_j for tracking.
5. Repeat Steps 4 to 5 until there's no better reply path can be found.
6. Perform the corresponding camera assignments and handoffs according to the set of camera action assignment $a = \{a_1, a_2, \dots, a_{N_C}\}$.
7. Repeat Steps 1 to 7 for every time instant. where δ is the improvement step.

It can be proved that given a small enough ε and a large number of iteration we will reach an optimal Nash equilibrium with an arbitrarily high probability. The proof involves implementing the resistance trees data structure [9] to describe games and is omitted because of the limited space. Our experimental results show that the convergence is reached pretty fast although there is no guarantee that within how many iterations one can get the optimal Nash equilibrium.

3. TWO OTHER TASK-ORIENTED APPROACHES

In this section, we will briefly introduce two task-oriented camera assignment approaches: the potential game approach in [7] and the constraint satisfaction problem approach in [8].

3.1. Camera Assignment as a Potential Game (PG)

In this approach, three utilities are concerned: *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. Unlike the weakly acyclic game, a potential game requires the person utility to be aligned with the global utility, i.e., 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, we must have

$$U_{P_i}(a_i', a_{-i}) < U_{P_i}(a_i'', a_{-i}) \Leftrightarrow U_g(a_i', a_{-i}) < U_g(a_i'', a_{-i}) \tag{3}$$

where a_{-i} stands for the assignments for persons other than P_i , i.e., $a_{-i} = (a_1, \dots, a_{i-1}, a_{i+1}, \dots, a_{N_P})$ and $a = (a_i, a_{-i})$. If the camera utility and global utility are defined the same as in the previous section, the person utility is defined as:

$$U_{P_i}(a) = \sum_{i=1}^{n_P} \sum_{m=1}^{3} w_m S_{im} - \sum_{\substack{i=1 \\ i \ne P_i}}^{n_P} \sum_{m=1}^{3} w_m S_{im} \tag{4}$$

where $i \ne P_i$ means that we exclude person P_i from the those who are being tracked by camera C_j.

The assignment profile is obtained by the bargaining mechanism [7]. In the k^{th} step, the assignment is made according to a set of probabilities

$$p_i(k) = [p_i^1(k), \dots, p_i^l(k), \dots, p_i^{nc}(k)]$$

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 $0 \le p_i^l(k) \le 1$, $l = 1, \dots, n_C$. At each bargaining step, $p_i^l(k)$ is calculated as

$$p_i^l(k) = \frac{e^{\frac{1}{\tau}\bar{U}^l_{P_i}(k)}}{e^{\frac{1}{\tau}\bar{U}^1_{P_i}(k)} + \dots + e^{\frac{1}{\tau}\bar{U}^{nc}_{P_i}(k)}} \tag{5}$$

where

$$\bar{U}^l_{P_i}(k+1) = \begin{cases} \bar{U}^l_{P_i}(k) + \frac{1}{p_i^l(k)}\left(U_{P_i}\big(a(k)\big) - \bar{U}^l_{P_i}(k), a_i(k) = A_i^l\right. \\ \bar{U}^l_{P_i}(k) \qquad\qquad\qquad\qquad , otherwise \end{cases} \tag{6}$$

$\bar{U}_{P_i}(k)$ is the estimated person utility: Before deciding the final assignment profile, we predict the person utility using the previous person's utility information in the bargaining steps. A_i is the set of available cameras for P_i. After several steps of calculation, the result of $p_i(k)$ tends to converges. Thus, we can finally get the stable solution.

3.2. Camera Assignment as a Constraint Satisfaction Problem (CSP)

The approach in [8] solves the camera assignment problem through solving the constraint satisfaction problem. If we allow one camera to track multiple persons but one person can only be tracked by one camera, for each camera C_j, we let all those persons that can be seen by this camera form a group g_j. For instance, if, in our case, the camera C_j can see person P_1 and P_2, then the domain of g_j, noted as Dom[g_j], is $\{\{P_1\}, \{P_2\}, \{P_1, P_2\}\}$. The constraint is set to be $d_i \cap d_j = \{\emptyset\}$, for $i \neq j$, where $d_i \in b_i \cup \emptyset$ is the camera assigned to track person P_i. b_i and b_j belong to Dom[g_j] and $i \neq j$. By doing so, we mean that the persons to be tracked are assigned to different cameras. We changed some of the notations in [8] so that the notations in this section are not in conflict with the notations used in the previous sections of this paper.

4. EXPERIMENTAL RESULTS

4.1. Criteria Used in Experiments

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, measured by the ratio of the number of pixels inside the bounding box of the person to that of the size of the image. Assume that λ is the threshold for best observation, i.e. when $r = \lambda$ this criterion reaches its peak value, where $r = \frac{\#\ of\ pixels\ inside\ the\ bounding\ box}{\#\ of\ pixels\ in\ the\ image\ plane}$.

$$S_{i1} = \begin{cases} \lambda r, \ when\ r < \frac{1}{\lambda} \\ \frac{1-r}{1-\lambda}, \ \ when\ r \ge \frac{1}{\lambda} \end{cases} \tag{10}$$

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

$$S_{i2} = \frac{\sqrt{(x-x_c)^2 + (y-y_c)^2}}{\frac{1}{2}\sqrt{{x_c}^2 + {y_c}^2}} \tag{11}$$

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 entire bounding box. We assume that the threshold for best frontal view is, i.e. when $R = \xi$ the view of the person is the best, where $R = \frac{\#\ of\ pixels\ on\ the\ face}{\#\ of\ pixels\ on\ the\ entire\ body}$.

$$S_{i3} = \begin{cases} \xi r, \ when\ R < \frac{1}{\xi} \\ \frac{1-R}{1-\xi}, \ \ when\ R \ge \frac{1}{\xi} \end{cases} \tag{12}$$

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

$$S_{i4} = \sum_{m=1}^{3} w_m Crt_{im} \tag{13}$$

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. In our experiments, we give value to the parameters empirically. $\lambda = \frac{1}{15}, \xi = \frac{1}{6}, w_1 = 0.2, w_2 = 0.1$ and $w_1 = 0.7$.

4.2. Experiments and Comparison Results

Experiments for 2 persons 3 cameras and 3 persons 4 cameras cases are carried out using the above criteria to maximize the system performance. We use the learning algorithm introduced in Section 2, the bargaining mechanism introduced in [7] and the *BestSlov* algorithm introduced in [8], respectively. In our experiments, an *error* is defined as either losing the track of a person or failing to select the camera with a frontal view whenever a frontal view is available.

- *The 2 person case*

For easier comparison, we use the same number of iterations for different approaches. In this case, we use 5 iterations for the proposed approach and the approach in [7]. We also use 5 backtracks in the Constraint Satisfaction Problem approach in [8].

In this simple case, all the three approaches can do the camera assignment well. In Table 1, we list the rates for successfully following the persons, successfully selecting the frontal views, the number of hand-offs that are taken place and the error rate for the three approaches, respectively. We can notice that since we prefer a frontal view to be selected, all these three approach can achieve a high rate of frontal view (whenever it is available). The two game theoretic approaches achieve slightly lower error rate than the constraint satisfaction problem approach. However, the drawback of the approach in [7] is that it requires modeling the camera assignment process as a potential game, which has more constraints than the weakly acyclic game. For instance, one has to find a way to design the person utility to be aligned with the global utility, which is not restricted in the weakly acyclic game. Thus, the utility function in our proposed approach is easier to design and can cover a larger scope of circumstances.

Table 1: Comparison results for the two person case.

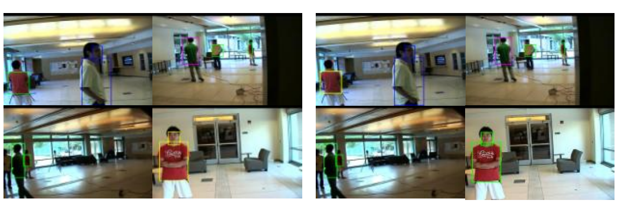
	Frontal view	Success following	No. of handoffs	Error rate
CSP	97.54%	87.98%	9	12.38%
PG	97.77%	83.02%	8	3.56%
WAG	98.02%	96.65%	8	2.79%

- *The 3 persons case*

In this case, the number of iterations or backtracks is fixed at 10 (for the reason similar to the previous case). As the number of cameras and persons goes up, the Constraint Satisfaction Problem based approach misses the “best” camera with a high probability. The high error rate for this approach is due to its computational cost. If we allow 27 backtracks for this case as shown in Table 2, the error rate for the Constraint Satisfaction Problem approach will be less than 5%. Figure 1 provides an example frame where the proposed approach selects the frontal view of the target person while the CSP approach fails because of lack of the number of backtracks.

Table 2: Comparison results for the three person case.

	Frontal view	Success following	No. of hand-offs	Error rate
CSP	89.75%	77.18%	21	26.66%
PG	94.00%	80.00%	17	9.02%
WAG	96.12%	89.50%	18	7.32%



WAP **CSP**

Figure 1: Example frame where the CSP approach fails to select the available frontal view while the proposed approach can. The green bounding boxes are for the stand-by (non-chosen) cameras.

- *The n persons case*

Figure 2 gives a comparison of number of iterations or backtracks for the three approaches for the case when the number of cameras is fixed to 3 and the number of persons goes up from 1 to 10.We can notice that as the situation becomes complicated, the convergence of the Constraint Satisfaction Problem approach becomes slower, i.e. the computational cost for this approach is the highest among all the three approaches that are compared in this paper.

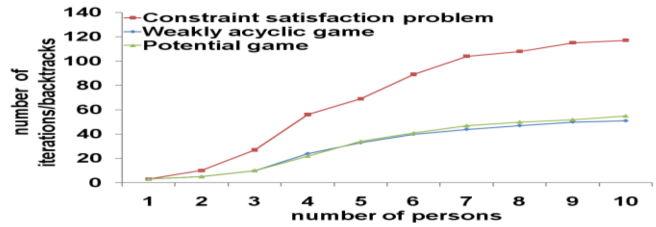


Figure 2: Comparison for the number of iteration or backtrack for the three approaches.

5. CONCLUSIONS

In this paper, we compared three approaches that can solve the camera assignment problem according to the tasks defined by the user. The experimental results show that the two game theory approaches are computationally more efficient than the Constraint Satisfaction Problem based approach. The merit of the proposed approach based on weakly acyclic game is that we formulate the camera assignment problem as a model that has fewer limitations and, therefore, it is applicable to a wide variety of situations. In the future, we will extend the application of the proposed approach to more complicated circumstances, such as in a network of Pan/Tilt/Zoom cameras. The weakly acyclic modeling is highly promising for active camera control by adding Pan/Tilt/Zoom to the camera action profile.

Acknowledgements: This work was supported in part by NSF grants 0551741, 0905671, ARO grant W911NF-07-1-0485 and ONR grant on Aware Building.

6. 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, Dec., 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”. *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 Int’l Workshop on Video Surveillance and Sensor Networks*, pp. 195-202, 2006.

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

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

[6] X. Zou and B. Bhanu. “Anomalous Activity Detection and Classification in a Distributed Camera Network”. *ICIP 2008*.

[7] Y. Li and B. Bhanu. “Utility-Based Dynamic Camera Assignment and Hand-off in a Video Network”. *ICDSC 2008*.

[8] F. Z. Qureshi and D. Terzopoulos. “Multi-Camera Control Through Constraint Satisfaction for Pedestrian Surveillance”. *AVSS 2008*, pp 211-218.

[9] J. R. Marden *et al.*. “Payoff based dynamics for multi-player weakly acyclic games”. *Proceedings of the 46th IEEE Conference on Decision and Control*, 2007.

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