Linear Production Game Solution to a PTZ Camera Network

YU-CHUN LAI[1], YU-MING LIANG[2], SHENG-WEN SHIH[3], CHENG-CHUNG LIN[1]
AND HONG-YUAN MARK LIAO[1,4]

[1]*Department of Computer Science and Information Engineering*
National Chiao Tung University
Hsinchu, 300 Taiwan
[2]*Department of Computer Science and Information Engineering*
Aletheia University
Danshui, 521 Taiwan
[3]*Department of Computer Science and Information Engineering*
National Chi Nan University
Nantou, 545 Taiwan
[4]*Institute of Information Science*
Academia Sinica
Taipei, 105 Taiwan

Reconfiguring the parameters of PTZ cameras in a surveillance system is a combinatorial optimization problem. Computing the optimal solution is very time consuming, and existing methods can only provide sub-optimal solutions. In this paper, we propose a non-linear objective function that better utilizes a camera network to track multiple targets. We also show that, by expanding the unknown parameters and imposing new constraints, the non-linear objective function can be converted into a linear production game (LPG) problem. Since an LPG yields an optimal solution that can be evaluated in polynomial time, the proposed method is efficient and accurate. The results of simulations and a real-world experiment demonstrate the proposed method's potential.

Keywords: visual surveillance, camera network, pan-tilt-zoom camera, linear production game, target tracking

1. INTRODUCTION

Intelligent video surveillance systems have been used for several years, and they are now widely deployed in important places, such as airports, all over the world. A single camera can provide useful information for event detection and target tracking [1, 2]; however, a surveillance system based on a camera network can reduce the number of blind spots and improve the system's reliability [3-9]. Camera networks are usually comprised of a heterogeneous collection of cameras, including panoramic cameras, fixed cameras, infrared cameras, and pan-tilt-zoom (PTZ) cameras. Among the different types of imaging devices, PTZ cameras are the most important components of an intelligent surveillance system because their field of view (FOV) can be changed in response to different task requirements. However, incorporating PTZ cameras into a surveillance system raises a challenging issue: How can the cameras be controlled and coordinated to accomplish a given task? Most surveillance tasks performed by PTZ cameras are related to three functions: tracking multiple targets, improving evidential quality and maximiz-

Received April 19, 2011; revised August 8, 2011; accepted September 23, 2011.
Communicated by Jen-Hui Chuang.

ing surveillance coverage.

Target tracking involves target detection as well as temporal and/or inter-view target correspondence matching. For example, Lim *et al.*'s approach [3] tracks the targets observed in FOVs and constructs a dynamic scene model containing the position, velocity, and view-dependent visibility of each target. The system tries to accomplish three objectives, namely, initial detection of moving objects, tracking of moving objects, and scheduling cameras to monitor activities. Cameras are assigned to tasks by solving a bipartite matching problem so that tasks are accomplished in order of priority. In the PTZ camera network developed by Ukita and Matsuyama [4], when the system detects a new target, nearby cameras that are idle are assigned to track the target. The system is simple and effective provided that the number of cameras is greater than the number of targets. Qureshi and Terzopoulos [5] introduced a multi-camera tracking system in which calibrated wide-FOV cameras are used to locate targets, and PTZ cameras are used to fixate on the located targets. The PTZ network operates according to heuristic rules designed to track targets cooperatively. Their method was further improved so that it can be applied to an uncalibrated multi-camera surveillance system connected with wireless communication network [10]. Since the wireless communication range is limited, Qureshi and Terzopoulos assumed that each camera can only communicate with its neighbors. Therefore, cameras within a communication range can share information of targets to be tracked. Adding/removing camera nodes can be accomplished very easily with their method. However, the flexibility is exchanged for security because, in general, a camera network sufficiently sharing information with a centralized server can outperform a camera network sharing information locally. In summary, a cooperative target tracking approach can utilize the camera resources efficiently and enable the cameras to support each other to recover the tracking when a tracking task fails.

While a solution to the target tracking problem can provide each target's trajectory, a surveillance system usually requires more information. For example, it is often necessary to capture the face of a human target or the license plate of a vehicle at a high resolution. These applications are related to improving evidential quality [7, 8]. In addition, PTZ cameras are frequently used to extend the coverage of the surveillance area by pan-tilt scanning. Piciarelli *et al.* [9] proposed an approach that reconfigures the pan-tilt-zoom parameters of all PTZ cameras based on the probability of observing an event in a specific location. Song *et al.* [6] applied game theory to maximize the surveillance coverage in their decentralized system, and adopted a sequential optimization strategy to achieve the Nash equilibrium [11]. Under this method, one PTZ camera is selected at random each time and its parameters are tuned, while those of the other cameras remain unchanged. After the Nash equilibrium is achieved, the cameras should cover the entire surveillance area at an acceptable resolution. When a human operator decides to track a specific target at a higher resolution, the target will be assigned to the most appropriate PTZ camera, which will then be excluded from the game. As a result, the remaining cameras have to adjust their parameters and try to maintain the maximum surveillance coverage. The game theoretic framework has two advantages: (1) it can be implemented easily; and (2) only a small amount of information needs to be exchanged. However, we remark that the Nash equilibrium is not necessarily an optimal solution. Moreover, tracking a specific target at a higher resolution is treated as an exceptional task that cannot be optimized by using the same game theory framework. Another potential problem of their method is that there is no mechanism to suppress investing too many resources (*i.e.*, cameras) in tracking a single target.

Reconfiguring PTZ cameras to achieve any of the three objectives is intrinsically a combinatorial optimization problem. Computing the optimal solution is very time consuming and, therefore, existing methods usually reduce the problem into a bipartite matching problem, which assigns tasks to cameras. However, the task assignment formulation does not fully utilize the camera network. For example, when the number of tasks is greater than the number of cameras, some tasks will have to be abandoned despite that a camera may accomplish multi-tasks at the same time.

In this paper, we propose an optimal and flexible solution to the PTZ network coordination problem. We show that the problem can be formulated as a linear production game (LPG). The LPG is about how a group of collaborative players utilizing their limited resources to create various products yielding the maximum payoff given that the price of each product is known. Players in the LPG of a PTZ network are the cameras. Each camera can control its FOV by selecting the PTZ parameters. Although the number of all PTZ combinations is very large, due to the limited speed of PTZ actuators, only a small set of new PTZ settings has to be considered. The new FOV corresponding to each new PTZ setting is the product of the game. Resources owned by each player (*i.e.*, camera) are the targets which are observable to the camera. The observability of a target to a camera is determined by checking whether the camera can select a feasible PTZ settings to observe the target. The price of a product (*i.e.*, the new FOV of a camera) is evaluated by examining the video quality of each target in the FOV. The goal of this cooperative game is to select a set of FOVs for the cameras to maximize the total payoff (video quality of the targets). The LPG is a special case of a linear programming problem. While a linear programming problem may be infeasible or unbounded [12], the LPG always has an optimal solution that can be evaluated in polynomial time. Therefore, many techniques, such as branch-and-bound and cutting-plane techniques [13], can be applied to solve the camera network reconfiguration problem.

The remainder of this paper is organized as follows. In section 2, we formulate the PTZ network problem; and in section 3, we describe the proposed LPG method. Section 4 details the results of simulations and a real-world experiment. Section 5 contains some concluding remarks.

2. PROBLEM FORMULATION

Suppose a surveillance system contains n calibrated PTZ cameras, each of which is controlled by a network-connected processor. In addition, a fixed (non-PTZ) camera in our system is seen as a PTZ camera with only one available FOV. Furthermore, let m be the number of targets detected in the surveillance area. Each detected target is represented by a status vector denoted by $\mathbf{g}_t^k = [\mathbf{b}_t^k, \mathbf{v}_t^k]$, where $\mathbf{b}_t^k$ and $\mathbf{v}_t^k$ are, respectively, the 3-D bounding box and the velocity of target k estimated at time t. The target's status $U_t = \{\mathbf{g}_t^k | k = 1, 2, \ldots, m\}$ and the static background constitute a dynamic scene model (targets' history positions and a top-view scene model) that can be used to predict the status of all the targets at time $t + 1$, expressed as $\hat{U}_{t+1} = \{\hat{\mathbf{g}}_{t+1}^k | k = 1, 2, \ldots, m\}$. The model is maintained by a central information processing node (a central server) that gathers information about the detected targets and the camera parameters from each camera node. It is assumed that the camera network has been calibrated and the homography (a point to point mapping matrix) between any two of the cameras is known so that the information about the detected target can be integrated. The central information processing node is also responsible for determining the optimal camera parameters.

2.1 Parameters to be Determined

Let ϕ^i denote the ith camera's FOV, which is controlled by the pan-tilt-zoom parameters of the camera. We assume that the relationship between the FOV and the parameters is known. Therefore, the problem of determining the optimal camera parameters is transformed into a problem of selecting the optimal FOV for each camera. Because of the limitation of the lens motor speed, a camera can only change its parameters locally in a short time. Hence, given each camera's current parameters, a set of feasible FOVs can be constructed and expressed as follows:

$$\Phi^i = \{\phi_j^i \,|\, j = 1, 2, \ldots, w_i\}, \tag{1}$$

for $i = 1, 2, \ldots, n$ where w_i is the number of feasible FOVs of the ith camera. The PTZ camera coordination problem is formulated as the following combinatorial optimization problem:

$$(\phi^{1*}, \ldots, \phi^{n*}) = \arg \max_{\phi^i \in \Phi^i, i=1,\ldots,n} Q(\phi^1, \ldots, \phi^n), \tag{2}$$

where $Q(\cdot)$: $\Phi^1 \times \ldots \times \Phi^n \mapsto \mathbb{R}$ is a function mapping $(\phi^1, \ldots, \phi^n)$ to a real quality value. In the next subsection, we explain how to assess the quality of a set of FOVs for different goals.

2.2 Quality Function of a Camera's FOV

Under the dynamic scene model, the locations of predicted targets can be computed for each camera's FOV. The predicted bounding box is defined as a region of interest (ROI). In a visual surveillance system, assessing the quality of an FOV usually involves the following two steps.

1. Determine whether the FOV includes some ROIs. An FOV without any ROIs should be evaluated as having the lowest quality.
2. Evaluate the dimensions (width and height) of each ROI. The resolution of the ROI should be sufficient to accomplish the given task. Aldrige and Gilbert [14] suggested different resolution requirements for different tasks. If a resolution is lower than the suggested value, a very low quality value should be assigned to it. Conversely, if the resolution is higher than the suggested value, the quality value should be upper bounded or reduced to induce camera zoom out for monitoring a larger area.

For most surveillance tasks, the quality of each camera's FOV can be evaluated individually and the total quality function, $Q(\phi^1, \ldots, \phi^n)$, can be simplified as follows:

$$Q(\phi^1, \ldots, \phi^n) = f\langle q_1, q_2, \ldots, q_n\rangle, \tag{3}$$

where $q_i = Q(\phi^i)$ for $i = 1, 2, \ldots, n$, and $f(\cdot)$: $\mathbb{R}^n \mapsto \mathbb{R}$ is a function that maps the n individual quality values to a total quality value. We discuss possible choices of $f(\cdot)$ later in this section. Furthermore, the quality function of each FOV, say ϕ^i, can be expressed as a function of the qualities of individual ROIs. Since the quality of each ROI can be evaluated independently, it is reasonable to compute the quality of an FOV as follows:

$$Q(\phi^i) = \sum_{k=1}^{m} Q_k(\phi^i), \tag{4}$$

where $Q_k(\phi^i) \triangleq Q(\hat{\mathbf{b}}_{t+1}^k; \phi^i, \hat{T}_{t+1} - \{\hat{\mathbf{g}}_{t+1}^k\})$ is the non-negative quality of observing the kth target with FOV ϕ^i. The value is zero when $\hat{\mathbf{b}}_{t+1}$ is not observable in FOV ϕ^i or it is completely occluded by other targets at T_{t+1}. The simplest form of $f(\cdot)$ in Eq. (3) is a linear summation function given by

$$Q_L\left(\phi^1, \ldots, \phi^n\right) = \sum_{i=1}^{n} Q\left(\phi^i\right). \tag{5}$$

However, since the qualities of a target in all views count toward the total quality function, maximizing (5) makes all the cameras pursue high video quality targets and ignore low quality ones.

To resolve the problem, we adopt the following non-linear quality function in this work:

$$Q_{NL}\left(\phi^1, \ldots, \phi^n\right) = \sum_{k=1}^{m} \max_i Q_k\left(\phi^i\right), \tag{6}$$

where only the maximum ROI quality of each target counts toward the total quality. Thus, the quality of a solution that favors a specific target will be lower than that of a solution that assigns the cameras to monitor different targets.

3. THE PROPOSED APPROACH

In this section, we show that the non-linear objective function (6) can be converted into a linear function by expanding the set of feasible solutions and imposing new constraints. Let T_j^i and $|T_j^i|$ denote the set of targets covered by FOV ϕ_j^i, *i.e.*, the jth feasible FOV of the ith camera (refer to Eq. (1)), and the number of targets in T_j^i respectively. The total number of subsets of T_j^i is $2^{|T_j^i|}$, and h is an index of the subset. For each subset $S_{j,h}^i \subset T_j^i, 1 \leq h \leq 2^{|T_j^i|}$, we can construct a virtual FOV that ignores any target not in $S_{j,h}^i$, *i.e.*, $Q_k(\phi_j^i) = 0$, for all $k \notin S_{j,h}^i$ (see Fig. 1).

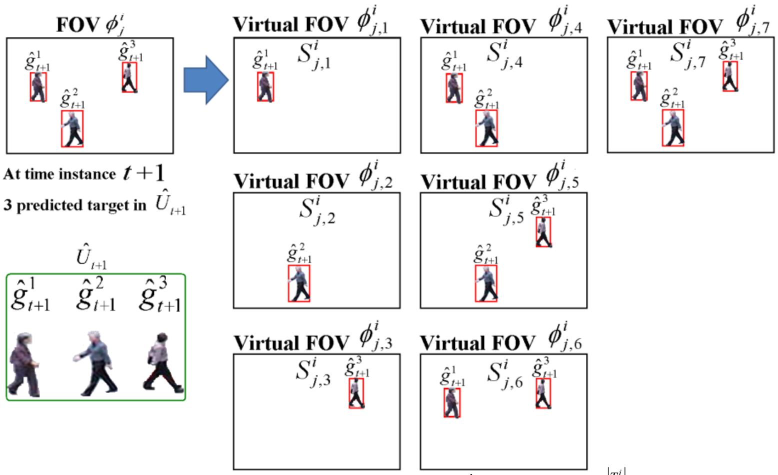


Fig. 1. The decomposition of an FOV containing $|T_j^i|$ targets into $2^{|T_j^i|} - 1$ virtual FOVs.

Notably, introducing virtual FOVs into the system increases the number of expanded feasible FOVs to $\sum_{i=1}^{n}\sum_{j=1}^{w_i} 2^{|T_j^i|} \leq n w_{max} 2^{|T_{max}|}$, where w_{max} and $|T_{max}|$ are the maximum number of feasible FOVs of a camera and the maximum number of targets in an FOV, respectively. From the complexity analysis, it is obvious that the computation load is linearly proportional to the number cameras and is exponentially proportional to the number of targets in an FOV. The exponential growth of the variables may lead to the scalability problem. However, since a PTZ camera is mainly used to acquire high definition images of targets by choosing a proper zoom setting, the number of targets observed by a PTZ camera is very limited. To give an impression about the typical number of targets covered by the FOV of an camera in different surveillance applications, we refer the suggested target sizes with respect to four different applications described by Aldrige and Gilbert [14]. According to the data shown in [14], the maximum number of targets in an FOV of a camera is less than 15 for the recognition and identification applications. The suggesting target size is estimated by the average human width-height ratio [15] (0.3442) and a 640 × 480 FOV. In practice, the maximum target number will be much smaller than this value and the solution can be computed very efficiently. The scalability problem emerges only when one uses too few PTZ cameras to observe too many targets. In that case, PTZ cameras are operating at the wide-angle (low resolution) mode, and the video content is less informative. To acquire useful surveillance videos, one should consider introducing more PTZ cameras into the network. Hence, although from the algorithmic point of view, the proposed method might suffer from scalability issues, in practice, the scalability issues can be ignored.

By replacing ϕ_j^i with the virtual FOVs, $\phi_{j,h}^i$, $h = 1, \ldots, 2^{|T_j^i|}$, the number of feasible FOVs to be assigned to the ith camera becomes $\sum_{j=1}^{w_i} 2^{|T_j^i|}$. To select the optimal FOVs, we define the binary variables $x_{j,h}^i \in$ B (B ≅ {0, 1}) to indicate whether the (j, h)th virtual FOV of the ith camera, *i.e.*, $\phi_{j,h}^i$, is selected. Therefore, the optimization problem in FOV selection can be rewritten as follows:

$$\max_{\mathbf{x}} \sum_{i=1}^{n} \sum_{j=1}^{w_i} x_{j,h}^{i} \sum_{k=1}^{m} Q_k\left(\phi_{j,h}^{i}\right), \tag{7}$$

subject to

$$\sum_{j=1}^{w_i} \sum_{h=1}^{2^{\left|T_j^i\right|}} x_{j,h}^{i} \leq 1, \tag{8}$$

For $i = 1, 2, \ldots, n$, and

$$\sum_{i=1}^{n} \sum_{j=1}^{w_i} \sum_{h=1}^{2^{\left|T_j^i\right|}} x_{j,h}^{i} o_{ijhk} \leq 1, \tag{9}$$

for $k = 1, 2, \ldots, m$, where $\mathbf{x} = \left[x_{1,1}^{1}, x_{1,2}^{1}, \cdots, x^{n}_{w_n, 2^{\left|T_{w_n}^n\right|}} \right]$ and the binary coefficient $o_{ijhk} \in \mathbb{B}$ indicates whether the kth target is observable in the virtual FOV $\phi_{j,h}^{i}$. The constraint specified in Eq. (8) ensures that each camera can only be assigned one FOV at a time, and Eq. (9) guarantees that the quality of a target can only be evaluated by a single FOV because the repeated target selection violates the constraint in Eq. (9) (the summation value is larger than 1.)

The relation between the solutions to Eqs. (6) and (7) can be derived by changing the summation order of Eq. (7) as follows:

$$\sum_{k=1}^{m} \left[\max_{\mathbf{x}} \sum_{i=1}^{n} \sum_{j=1}^{w_i} x_{j,h}^{i} Q_k\left(\phi_{j,h}^{i}\right) \right] = \sum_{k=1}^{m} q_k^{*}, \tag{10}$$

where q_k^{*} is the quality of the kth target evaluated by one of the optimal FOVs.

The objective function and the constraints given in Eqs. (10), (8) and (9) form a linear programming problem. Since a linear programming problem may be infeasible or unbounded [12], it is important to show that the above formulation yields an optimal solution that can be solved efficiently. In the following subsection, we show that our formulation can be regarded as an LPG [16]. The LPG is a special case of linear programming problems which has the following nice properties

1. The optimal solution of an LPG always exists, and
2. The solution can be computed in polynomial time by solving its dual problem [16, 17].

3.1 Linear Production Game (LPG) Problem

An LPG is a special case of a cooperative game that is usually denoted as (N, v), where N is a set of n players, namely $\{1, 2, \ldots, n\}$, and v: $2^N \to \mathbb{R}$ is a payoff function that maps a coalition to a payoff value [18]. Specifically, the payoff of an empty set is zero. A coalition, say S, is a subset of all players, *i.e.*, $S \subseteq N$. The LPG described in [16] involves maximizing the payoff of n cooperative players who own m different types of resources. The total number of the resources type k in a coalition S are denoted as $b_k(S)$, $k = 1, 2, \ldots, m$, and those resources can produce p different products. A resource-product matrix $\mathbf{A}$ is used to specify the relationships between the amounts of the different types of resources required to produce each product. The (i, j)th entry of the resource-product matrix, denoted by a_{ij}, describes the amount of the ith resource required to produce the jth product, which is non-negative. Then, each product can be sold at a different price given by $\mathbf{c} = (c_1, c_2, \ldots, c_p)$. In this game, the maximum payoff can be obtained by solving the following linear programming problem

$$v(S) = \max_{\mathbf{x}} \mathbf{c}^{\mathrm{T}}\mathbf{x}, \tag{11}$$

subject to $\mathbf{Ax} \leq b_k(S)$ and $x_i \geq 0$ for all i.

The above optimization problem indicates that all players in the coalition S should combine their resources to produce a set of products that maximize the revenue $v(S)$. Let $\mathbf{y} = [y_1 \ldots y_n]^{\mathrm{T}}$ denote the payoff vector, where $y_i = c_i x_i$, $i = 1, 2, \ldots, n$. In cooperative game theory, an imputation describing a payoff vector is defined as follows,

$$\{\mathbf{y} \mid \mathbf{y} = [y_1, y_2, \ldots, y_n]^{\mathrm{T}}, \sum_{i \in N} y_i = v(N), y_i \geq v(\{i\} \forall i \in N)\}. \tag{12}$$

Eq. (12) indicates that each player benefits by cooperating in a game because the resulting payoff, *i.e.*, y_i, of the ith player is not less than the payoff derived by playing alone, *i.e.*, $v(\{i\})$. $v(\{i\})$ means the maximum payoff value obtained by the ith player's resource. Moreover, in the cooperative game $v(N)$, the sum of the payoffs to all the players is equal to the total payoff. The solution concept of a cooperative game is called the core. It consists of one or more imputations that satisfy the following condition: all the payoffs of a subset in the imputation must be maximized. In the payoff vector, none of the players has a motive to leave the coalition because doing so would reduce his/her payoff. The core is defined as follows:

$$\mathbf{C} = \left\{ \mathbf{y} \mid \mathbf{y} = [y_1, y_2, \ldots, y_n]^{\mathrm{T}}, \sum_{i \in N} y_i = v(N), \sum_{i \in S} y_i \geq v(S) \forall S \subseteq N \right\} \tag{13}$$

Notably, the coefficients in the LPG equations are all positive. As the coefficients in the objective function (7) and in the constraints (8) and (9) are also all non-negative, the optimization problem can be easily mapped to an LPG problem. The role mapping between the camera network reconfiguration problem and the LPG is listed as follows:

1. The players are the cameras.
2. The resources of a player are the targets covered by a camera at the next time instance.
3. The products are the selected FOVs for the cameras.
4. The price of each product (*i.e.*, an FOV) is the quality of each selected FOV.

This formulation is similar to a bounty hunter game in which the cameras act as hunters who use FOVs to "hunt" targets. Different hunters will receive different rewards (*i.e.*, video quality) when obtaining a target. This is a cooperative game because all the hunters tend to let the hunter who may receive the greatest reward to catch a target. The existence of a core of the game ensures that all players will obtain the maximum payoff.

In this work, we adopt the branch-and-cut algorithm [19] to compute an optimal integer solution. Since an optimal solution guarantees that Eq. (10) is an upper bound of Eq. (6), the LPG solution is equivalent to the optimal solution of the non-linear objective function (6).

4. SIMULATIONS AND A REAL WORLD EXPERIMENT

To compare the performances of different approaches, we conducted three computer simulations and a real experiment. Throughout the four experiments, the following simple quality function is adopted.

$$Q_k(\phi^i_{j,h}) = \begin{cases} 1+\lambda z^i_{j,h}, & \text{if target } k \text{ is cover by } \phi^i_{j,h} \\ 0, & \text{otherwise,} \end{cases} \tag{14}$$

where $\lambda = 0.01$ in the experiments and $z^i_{j,h} \in [0, 1]$ is the zoom level of the FOV $\phi^i_{j,h}$. When the lens is fully zoomed out, $z^i_{j,h} = 0$. Conversely, $z^i_{j,h} = 1$ indicates that the lens is fully zoomed in. Therefore, $\lambda z^i_{j,h} \leq 0.01$. The quality function defined in Eq. (14) indicates that our goal is to control the PTZ cameras to capture as many as possible the targets appearing at the area under surveillance while favoring high zoom level camera settings. The goal can be applied to real-world applications. For example, the most common application is to record all targets' trajectories in the scene. Then, when an event occurred, the recorded trajectories can be utilized to trace a target (for example, a criminal) and to reconstruct his escaping route. For simplicity, we assume that the image resolution of any static camera included in the surveillance network is too low that the image quality function always returns zero. Hence, the optimal solution will be a control strategy that uses the PTZ cameras to capture the highest number of targets.

In the first two experiments, we simulated a surveillance system containing three virtual PTZ cameras, a virtual static camera covering the entire area under surveillance, and several virtual pedestrians with different paths and movement speeds. In total, we produced five pedestrian sets each containing 15 moving objects. Fig. 2 shows the simulation environment in which the total coverage of each PTZ is depicted as a fan-shaped region and a moving target is represented as a trajectory with time ordering.

We compared the performance of the proposed non-linear objective function LPG (NOF-LPG) method with that of the following four methods: the linear sum method, *i.e.*, the linear objective function LPG (LOF-LPG), which maximizes Eq. (5); Song *et al.*'s method (SONG) [6]; Lim *et al.*'s method (LIM) [3]; and the exhaustive search (ES) method. To find an optimal solution, the ES algorithm tests all possible combinations $\prod_{i=1}^{n} w_i$ of the feasible FOVs (refer to Eq. (1)) of all the cameras. The small-scale tests enable us computing an ES solution as a reference to compare different approaches. The simulated system was implemented in C++ on a PC with an Intel Core 2 DUO E8400 3.00Ghz CPU.

The first simulation compares the efficiency and accuracy of the five algorithms on the five pedestrian sets. The average computation times using NOF-LPG, LOF-LPG, SONG, LIM, and ES are shown in Table 1. The computational load of NOF-LPG is exponentially proportional to the maximum number of targets in an FOV, *i.e.*, $|T_j^i|$ (refer to section 3) but the number of targets in an FOV is normally small. Therefore, the efficiency of NOF-LPG is acceptable. In the simulation, the maximum number of targets in an FOV is five, so the computation time of the proposed method is satisfactory.

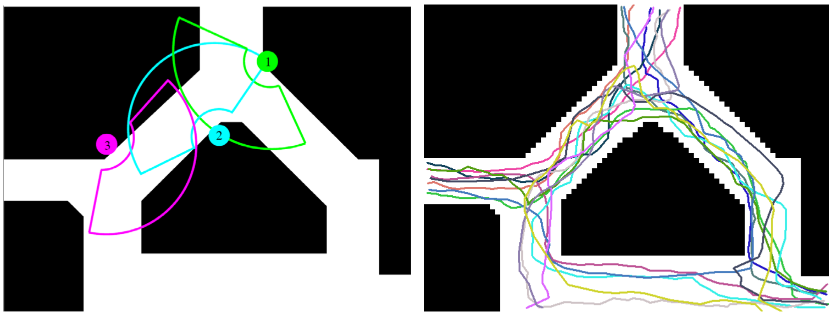


Fig. 2. The simulated surveillance environment with three PTZ cameras and fifteen moving targets in Pedestrian set 1.

Table 1. The average computation time required for one iteration by the NOF-LPG, LOG-LPG, SONG, LIM, and ES methods.

	Average computation time in the first simulation
NOF-LPG	334.14 ms
LOF-LPG	247.76 ms
LIM	236.33 ms
SONG	245.01 ms
ES	2330.37 ms

Fig. 3 shows the true number of moving objects in pedestrian set 1 and the number observed by ES. Note that the cameras may not be able to observe all the moving objects simultaneously because of the limitations of their FOVs. Therefore, the numbers of observed targets derived by the ES algorithm are used as a benchmark. In the simulation, the positions of all pedestrians are provided by the global-view static camera. Table 2 details their performances on the five pedestrian sets, where all data are normalized by the corresponding reference values (the result of ES). The NOF-LPG and LOF-LPG methods outperform the other two methods. The results show that the number of targets observed using NOF-LPG is equal to that of the ES algorithm. Furthermore, the LOF-LPG method outperforms the SONG and LIM methods because it utilizes a centralized optimization technique.

The proposed NOF-LPG method outperforms the other three methods because it adopts a better objective function.

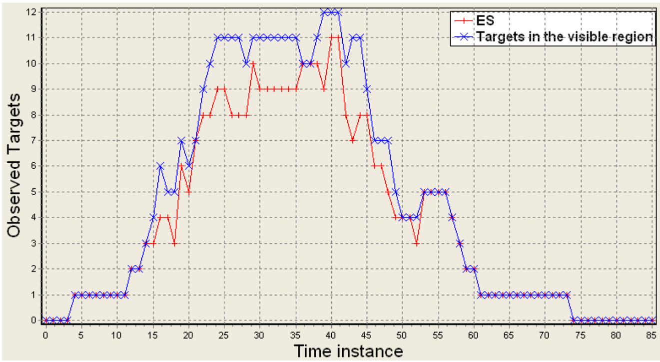
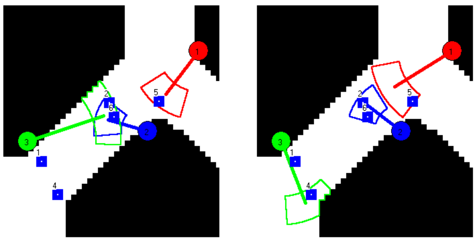


Fig. 3. The true number of moving objects in Pedestrian set 1 and the number observed by the exhaustive search (ES) method.

Table 2. The percentages of targets observed by the compared methods (normalized by the result of exhaustive search).

	Pedestrian set 1	Pedestrian set 2	Pedestrian set 3	Pedestrian set 4	Pedestrian set 5
NOF-LPG	100.0%	100.0%	100.0%	100.0%	100.0%
LOF-LPG	96.4%	93.7%	93.9%	93.6%	93.2%
LIM	75.6%	71.7%	84.6%	77.9%	77.0%
SONG	73.4%	76.0%	66.3%	72.4%	68.4%

Time instance 21



LOF-LPG NOF-LPG

Fig. 4. The snapshot of NOF-LPG and LOF-LPG at time instance 21.

As mentioned in section 2.2 that the linear summation function (5) encourages all cameras to pursue high video quality targets without a mechanism to suppress too many cameras focusing on the same group of targets. Therefore, the LOF-LPG method may lose track of some targets. Fig. 4 shows an example of this effect found in the first simulation at time instance 21. This figure shows that target 4 is ignored by the LOF-LPG algorithm because the sum of the quality values of targets 2 and 6 is greater than that of target 4. Conversely, the NOF-LPG appropriately controls cameras' FOVs to obtained more targets.

We also conducted an experiment using real-world data to evaluate the performance of the four methods. A camera network comprised of two PTZ cameras (AXIS AX-215) and one fixed camera (D-link DCS-3220G) was deployed in an outdoor environment. There are quite a few automatic methods to estimate the image homographies, such as [20] and [21]. However, since estimating a homography only requires four pairs of correspondence points, it is not difficult to manually construct the correspondence points. Therefore, in the experiment, all cameras were calibrated by selecting landmarks manually and the homography map was computed by using functions in OpenCV [22].

The video resolution of each FOV was 352 × 240. At the system setup stage, 24 predetermined FOVs were assigned to each PTZ camera. To calibrate the homographies of the predetermined FOVs, the images corresponding to the 24 FOVs were acquired and used to build a panoramic image of the PTZ camera with Autostitch [23]. Therefore, the homography between any of the 24 FOVs and the corresponding panoramic image is known. At present, the 24 FOVs only contain two zoom levels because, if the FOV is too narrow, the image features would be insufficient (feature points < 4) to estimate the image homography reliably. The panoramic images of the two PTZ cameras and the images of the fixed camera are related by registering them to a top-view aerial image. Fig. 5 (a) shows the set of 24 acquired images and the registered panoramic image; and Fig. 5 (b) shows the visual coverage regions of the three cameras overlaid on a top-view image. Therefore, the homography between any two cameras is known.

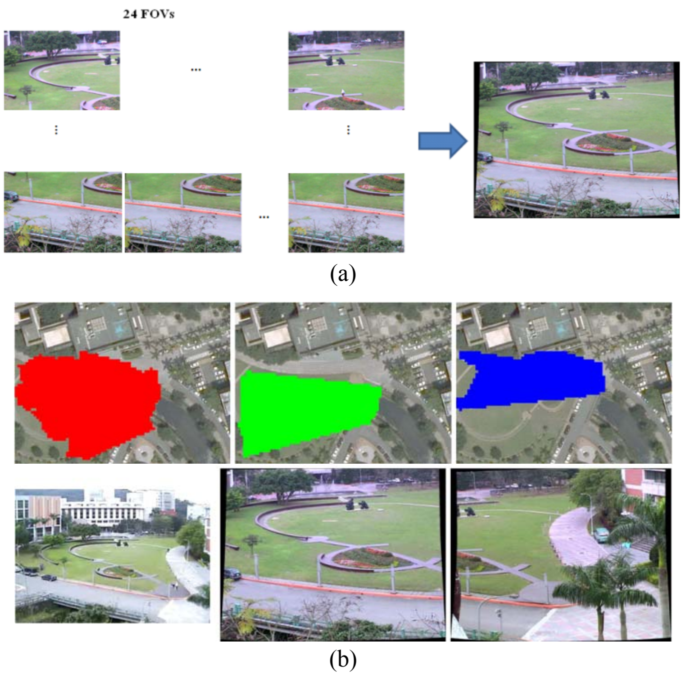


Fig. 5. (a) A panoramic image derived by integrating 24 images; (b) the coverage regions of the three cameras overlaid on a top view image and the panoramic image of the three cameras.

The video from the fixed camera is used for target detection and prediction; and Yoo and Park's difference-based approach [24] is utilized to detect moving targets. Although the approach in [24] may fail to detect semi-stationary targets, it is efficient and robust to lighting variations; thus, it is very suitable for an outdoor real time system. A simple constant-velocity motion model is used to predict the locations of targets.

Since it is impossible to control the PTZ cameras by using the four compared methods simultaneously, we tested them one by one. For comparison, each algorithm was tested for 20 minutes (about 5400 frames) on a cloudy and windy day. At each iteration, the system required about 0.8 seconds to estimate the parameters of the constant-velocity model of each moving target in order to predict its next position. Then, new parameters were determined and sent to the PTZ cameras. Usually, the cameras took 0.6 seconds to move to the assigned FOV. Finally, the targets observed by the FOVs were counted to evaluate the performance. Fig. 6 shows the NOF-LPG results at three time instances. In video frame 1470, several groups of people are detected and the two PTZ cameras are instructed to observe different groups of targets for maximizing the observed target number. Video frames 3358 and 3444 exemplify a target hand-over between cameras, where the group observed by camera 2 in frame 3358 is passed to camera 3 in frame 3444. As in the simulations, the exhaustive search method is applied to obtain reference data. However, since it is impractical to test all combinations of the cameras' FOVs when determining the optimal solution in a real environment, the exhaustive search is based on the video recorded by the fixed camera. Each feasible FOV of the PTZ cameras is mapped to a region in the image of the fixed camera. Then, all combinations of the mapped regions are utilized to search for the maximum number of observed targets as reference data. Table 3 details the performances of the four algorithms in the real-world experiment.

Fig. 6. The tracking results of frames 1470, 3358, and 3444 using NOF-LPG.

Table 3. The percentages of targets observed by the compared methods in a real environment.

	Performance (normalized by the result of exhaustive search in the fixed camera)
NOF-LPG	71.0%
LOF-LPG	67.3%
LIM	60.2%
SONG	62.3%

Although the performance of each algorithm is affected by shadows, wind, calibration errors, target detection errors, and prediction errors, the results demonstrate that NOF-LPG and LOF-LPG still outperform LIM and SONG.

5. CONCLUDING REMARKS

In this paper, we have considered the process used to assess the quality of a set of FOVs and proposed a non-linear objective function to reduce the number of unattended targets in a surveillance region. When the quality of each FOV is evaluated individually, it is easy to derive the objective functions of different PTZ network problems. The proposed approach provides an optimal solution for the PTZ network coordination problem. We have also shown that the non-linear optimization problem can be transformed into a linear production game problem that is guaranteed to yield an optimal solution. The optimal solution of LPG can be computed in polynomial time so our approach is efficient. The branch-and-cut method is adopted to solve the PTZ parameters selection problem. Computer simulations and a real-world experiment were performed to evaluate the proposed method in a multi-target surveillance environment. The results show that the method achieved the highest tracking rate among the compared methods. In future work, we will address the system efficiency to improve the frame rates of the real-world experiment. Since the system preprocessing such as target detection and target prediction can be performed independently, distributing the preprocessing to camera nodes is an available way to descrease the computational loads of the central server.

REFERENCES

1. I. Haritaoglu, D. Harwood, and L. S. Davis, "W4: real-time surveillance of people and their activities," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 22, 2000, pp. 809-830.
2. O. Javed and M. Shah, "Tracking and object classification for automated surveillance," in *Proceedings of the European Conference on Computer Vision*, 2002, Vol. 2423, pp. 343-357.
3. S.-N. Lim, L. S. Davis, and A. Elgammal, "A scalable image-based multi-camera visual surveillance system," in *Proceedings of the IEEE Conference on Advanced Video and Signal Based Surveillance*, 2003, pp. 205-212.
4. N. Ukita and T. Matsuyama, "Real-time cooperative multi-target tracking by communicating active vision agents," *Computer Vision and Image Understanding*, Vol.

97, 2005, pp. 137-179.
5. F. Z. Qureshi and D. Terzopoulos, "Surveillance in virtual reality: System design and multi-camera control," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2007, pp. 1-8.
6. B. Song, C. Soto, A. K. Roy-Chowdhury, and J. A. Farrell, "Decentralized camera network control using game theory," in *Proceedings of the ACM/IEEE International Conference on Distributed Smart Cameras*, 2008, pp. 1-8.
7. A. D. Bagdanov, A. D. Bimbo, and W. Nunziati, "Improving evidential quality of surveillance imagery through active face tracking," in *Proceedings of the International Conference on Pattern Recognition*, Vol. 3, 2006, pp. 1200-1203.
8. N. Bellotto, E. Sommerlade, B. Benfold, C. Bibby, I. Reid, D. Roth, L. V. Gool, C. Fernandez, and J. Gonzalez, "A distributed camera system for multi-resolution surveillance," in *Proceedings of the ACM/IEEE International Conference on Distributed Smart Cameras*, 2009, pp. 1-8.
9. C. Piciarelli, C. Micheloni, and G. L. Foresti, "PTZ camera network reconfiguration," in *Proceedings of ACM/IEEE International Conference on Distributed Smart Cameras*, 2009, pp. 1-7.
10. F. Z. Qureshi and D. Terzopoulos, "Smart camera networks in virtual reality," in *Proceedings of the IEEE*, Vol. 96, 2008, pp. 1640-1656.
11. E. Rasmusen, *Games and Information – An Introduction to Game Theory*, Wiley-blackwell, USA, 1991.
12. R. J. Vanderbei, *Linear Programming: Foundations and Extensions*, International Series in Operations Research & Management Science, Springer-Verlag, 2008.
13. S. P. Bradley, A. C. Hax, and T. L. Magnanti, *Applied Mathematical Programming*, Addison-Wesley, USA, 1977.
14. J. Aldrige and C. Gilbert, *Performance Testing CCTV Perimeter Surveillance Systems*, Number 14-95, White Crescent Press, Home Office Police Scientific Development Branch, Woodcock Hill, Sandridge, St Albans, Herfordshire, UK, 1996.
15. R. Lackore, *Firefighter Anthropometric Data White Paper*, FAMA Technical Committee Chassis Subcommittee, 2007.
16. G. Owen, "On the core of linear production games," *Mathematical Programming*, Vol. 9, 1975, pp. 358-370.
17. J. Derks and J. Kuipers, "On the core of routing games," *International Journal of Game Theory*, Vol. 26, 1997, pp. 193-205.
18. J. M. Bilbao, *Cooperative Games on Combinatorial Structures*, Kluwer Academic Publishers, Netherlands, 2000.
19. J. Mitchell, *Branch-and-Cut Algorithms for Combinatorial Optimization Problems in Handbook of Applied Optimization*, Oxford University Press, UK, 2002.
20. S. Calderara, A. Prati, and R. Cucchiara, "Hecol: Homography and epipolar-based consistent labeling for outdoor park surveillance," *Computer Vision and Image Understanding*, Vol. 111, 2008, pp. 21-42.
21. S. N. Sinha and M. Pollefeys, "Towards calibrating a pan-tilt-zoom camera network," in *Proceedings of the 5th Workshop on Omnidirectional Vision, Camera Networks and Non-classical Cameras*, 2004.
22. "Opencv library," http://opencv.willowgarage.com/wiki/.
23. M. Brown and D. Lowe, "Automatic panoramic image stitching using invariant fea-

tures," *International Journal of Computer Vision*, Vol. 74, 2007, pp. 59-73.
24. Y. Yoo and T.-S. Park, "A moving object detection algorithm for smart cameras," in *Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops*, 2008, pp. 1-8.

Yu-Chun Lai (賴育駿) received the B.S. degree in information science from Chinese Culture University, Taiwan, in 2002, and the M.S. degree in computer science and information engineering from Fu Jen University, Taiwan, in 2004. He is currently pursuing the Ph.D. degree in the Department of Computer Science, National Chiao-Tung University, Hsinchu, Taiwan. His current research interests include video analysis, video surveillance system, and image processing.

Yu-Ming Liang (梁祐銘) received the B.S. and M.S. degrees in Information and Computer Education from National Taiwan Normal University, Taiwan in 1999 and 2002, respectively, and the Ph.D. degree in Institute of Computer Science and Engineering from National Chiao Tung University, Taiwan in 2009. From January 2009 to Jan. 2010, he was a Postdoctoral Fellow in the Institute of Information Science, Academia Sinica, Taiwan. Since February 2010, he joined the Department of Computer Science and Information Engineering, Aletheia University, Taiwan, as an Assistant Professor. His research interests include computer vision, pattern recognition, and multimedia signal processing.

Sheng-Wen Shih (石勝文) received the M.S. and Ph.D. degrees in electrical engineering from the National Taiwan University, Taiwan, in 1990 and 1996, respectively. From January to July 1997, he was a Postdoctoral Associate in the Image Formation and Processing Laboratory, Beckman Institute, University of Illinois at Ubana-Champaign, Urbana. In August 1997, he joined the Department of Computer Science and Information Engineering, National Chi Nan University, Taiwan, Nantou, as an Assistant Professor, and became a Professor in 2010. His research interests include computer vision, biometrics, and human-computer interaction.

Cheng-Chung Lin (林正中) received his B.S. degree in Control Engineering from National Chiao-Tung University, Taiwan, in 1979 and M.S. degree of Computer Engineering from NCTU in 1985. He worked as a digital circuits/systems design engineer in the R&D Department, TECO Inc., Taiwan, during 1979 and 1981 and then worked at NCTU until 1990. He received his Ph.D. degree from the EECS Department of Northwestern University in 1997. He is now an Associate Professor of the Department of Computer Science, NCTU, Taiwan. His research interests include computer architectures, computer graphics, computer vision and face recognition.

Hong-Yuan Mark Liao (廖弘源) received a BS degree in physics from National Tsing-Hua University, Hsin-Chu, Taiwan, in 1981, and an M.S. and Ph.D. degree in electrical engineering from Northwestern University, Evanston, IL, in 1985 and 1990, respectively.

He is also jointly appointed as a Professor of the Computer Science and Information Engineering Department of National Chiao Tung University and the Electrical Engineering Department of National Cheng Kung University. Since February 2009, he has been jointly appointed as the Multimedia Information Chair Professor of National Chung Hsing University. Since August 2010, he has been appointed as Adjunct Chair Professor of Chung Yuan Christian University.

His current research interests include multimedia signal processing, video-based surveillance systems, content-based multimedia retrieval, and multimedia protection. He is on the editorial boards of the IEEE Transactions on Image Processing, the IEEE Transactions on Information Forensics and Security, and the IEEE Signal Processing Magazine. He was an Associate Editor of IEEE Transactions on Multimedia during 1998-2001.

Dr. Liao was the recipient of the Young Investigators' award from Academia Sinica in 1998. He received a Distinguished research award from the National Science Council of Taiwan in 2003 and 2010, respectively; a National Invention Award of Taiwan in 2004, a Distinguished Scholar Research Award from the National Science Council of Taiwan in 2008, and 2010 Academia Sinica Investigator Award. In June 2004, he served as the conference co-chair of the 5th International Conference on Multimedia and Exposition (ICME 2004) and technical co-chair of 2007 ICME. Dr. Liao served as the President of the Image Processing and Pattern Recognition Society of Taiwan during 2006-2008.