AUCTION PROTOCOL FOR CAMERA ACTIVE CONTROL

Yiming Li, Bir Bhanu and Wei Lin

Center for Research in Intelligent Systems, University of California, Riverside, CA

ABSTRACT

In this paper, we apply the auction-based theories in economics to camera networks. We develop a set of auction protocols to do camera active control (pan/tilt/zoom) intelligently. Unlike the economic auction, the bid price in our case is formulated to have a vector representation, such that when a camera is available to follow multiple objects, we consider the "willingness" of this camera to track a particular object. Most of the computation is decentralized by computing the bid price locally while the final decision is made by a virtual auctioneer based on all the available bids, which is analogous to a real auction in economics. Thus, we can take the advantages of distributed/centralized computation and avoid their pitfalls. The experimental results show that the proposed approach is effective and efficient for dynamically active control based on user defined performance metrics.

Index Terms— auction, active control, camera network

1. INTRODUCTION

The problem of efficiently applying camera active control (pan/tilt/zoom) has risen to the forefront of the video sensor networks. It can help to minimize the number of cameras in a video network or enlarge the covering area. How to intelligently pan/tilt/zoom the cameras in a video network to follow multiple targets is discussed in this paper. There are several related questions. When active control of cameras is available, how can we know in advance whether it is better to pan or tilt a camera to follow a person, who is originally not in its field-of-view (FOV), or to use a currently available camera to follow that person? When a camera can "see" more than one person in its monitored range (all the areas that can be possibly covered by a camera by panning or tilting, even if it may not be fully covered for the current setting) how does a camera decide which one to follow? All these questions are to be addressed in this paper.

There is a large amount of work done in the field of multi-camera multi-person tracking [1, 2]. Only a few works consider camera active control [3, 4, 5]. In [3] and [4], the authors only show results with single camera. In [5], the author focuses on the calibration of PTZ cameras online, rather than following the targets in the system. Most of the work does not focus on controlling active cameras in a video network to follow objects. Also, most of the work does not consider potentially available cameras, which means the camera may be available by active control. This paper aims to consider all possibilities, including the FOV after panning/tilting a camera, and find the best available camera to follow an object. Some existing approaches use greedy algorithms to find solutions for camera assignment, which prevents them from achieving the global optimum. In order to avoid sticking in local optima, we deploy the auction theory in economics.

Auction-based technique shows its effectiveness in solving many problems in multi-agent systems. For example, auction-based mechanism is established in [6] for computational grids. [7] uses the auction method for dynamic task allocation for groups of failure-prone autonomous robots. [8] proposes an opportunistic optimization approach for auction-based multirobot control. Leaders are used to do optimization within subgroups. Chen *et al.* [9] achieve single target tracking in wireless networks by deploying auction-based coalition. Qureshi and Terzopoulos [10] deploy an auctioning process to form groups of cameras in a virtual network. As compared to this paper they do not consider active control of cameras in the auctioning process. Moreover, they do not test the auctioning process in a real-world physical camera network.

In this paper, we model the process of selecting cameras to follow multiple objects in a camera network as the process of an economic *auction* [11]. In our refined model, there is a virtual auctioneer holding an auction for each object to be followed and all the potential cameras *bidding* for it. We develop a set of bidding protocols to involve camera active control. Unlike in [10], we make the bid price from each bidder a vector to consider its potential ability to follow an object by being panned or tilted and, thus, integrate camera active control into the auction process. By doing so, we benefit from the auction mechanism for distributed computation and consider the "willingness" of buyers (cameras). We choose the top bid to make the camera with the highest potential.

2. TECHNICAL APPROACH

2.1. Auction Protocols

2.11. *Problem formulation and notations:* An *auction* is the process of selling an item from the auctioneer to many potential buyers, i.e., bidders. Typically, in the auction, the bidders first offer their prices, bid [10, 11]. If the bidders

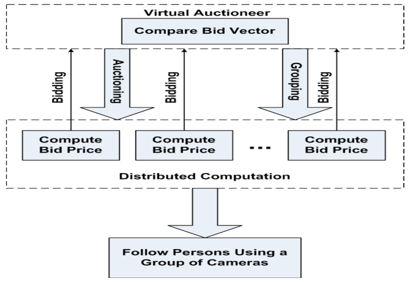


Figure 1: Overview of the auction-based approach.

bid for profitable trades only, we say that they are *rational*. Then, the auctioneer collects the bids information, and decides who wins the item and how much the winner has to pay.

The goal of the proposed approach is to involve camera active control by developing proper auction-bid mechanisms and form groups of cameras dynamically to follow multiple objects in the camera network. We want to select the camera with the best *quality of view (QOV)* for an object, based on pre-defined metrics, to follow the object. This camera may be the one that currently can "see" the object, or the one that may have a high QOV by panning or tilting to somewhere else.

A virtual auctioneer (a component that is not a real device like a camera, but is manipulated by the program) holds an auction for each of the objects in the system. All the potentially available cameras are modeled as potential buyers for the object. There is a set of metrics according to which the cameras will evaluate their willingness to follow the object or not and if they decide to follow, how much bid price they will provide. The auctioneer collects all this information and finally makes a decision, i.e., which camera to use to follow the object. This process is overviewed in Figure 1.

Before describing the detailed approach, we first clarify some assumptions made in our system:

1. Homographies are calculated and the cameras' heights are known, so that we know the coordinate conversion between different camera images.
2. The camera's focal length is set to a fixed number such that the angle of view (the largest angle that a camera can cover without any active control) is 51.2°. Each camera has 8 overlapping pre-defined pan settings to seamlessly cover 360 degrees. Also, there are three tilt settings, up 5°, down 5° (or -5°) and no tilt (0°). So, there are 24 settings for each cameras. We will call these 24 settings for Camera C_j as $\boldsymbol{l} = \{l_j^1, l_j^2, \dots, l_j^{24}\}$ where l_j^1 is the current location of Camera C_j.
3. The cameras are rational and honest, i.e. they calculate their bid price solely based on the pre-defined metrics and they will only do the profitable trades.
4. There is no communication error.
5. There is no communication congestion.

Some key notations that will be used in this paper are given in Table 1.

Table 1: Symbols and Notations.

Symbols	Notations
P_i	Person *i*
C_j	Camera *j*
n_C	The number of cameras that can "see" P_i
$\boldsymbol{l}$	Camera setting vector
l_j^k	The k^{th} setting of C_j
B_{ij}	Bid price sent from camera C_j for person P_i
$\boldsymbol{b_{ij}}$	Bid vector from C_j for P_i
b_{ij}^k	Intermediate bid from C_j for P_i at the setting l_j^k
M_{ijm}	The m^{th} metric score for P_i in C_j
w_m	Weights for different metrics
α_k	Weight on k^{th} dimension in bid price function
λ	Elasticity of substitution between different dimensions in bid price function

2.12. ***Auction Protocol:*** We develop a set of auction protocols to select the best available or potentially available camera to follow an object, which is described as follows:

1. Auction announcement. A virtual agent (program running on a central server) holds an auction for each object to be tracked. An auction message is broadcast to the whole network. The message includes information such as the location of an object and camera IDs of those cameras which are currently avaibale to follow it. Note that we will initialize the location of the object by a motion detection module. The camera that first "sees" the object will be initialized to follow this object. The object's location is initialized as the centroid location in the camera's image. After that, the camera used to follow this object is decided as the one with the highest bid price and the object's centroid in this camera will be broadcast.

2. Compute bid price. The overall bid price B_{ij}, which is from camera C_j for person P_i, is decided by a 24-dimensional bid vector, $\boldsymbol{b_{ij}} = \{b_{ij}^1, b_{ij}^2, \dots, b_{ij}^k, \dots, b_{ij}^{24}\}, k \in [1,24]$. b_{ij}^k stands for the *intermediate bid* that the camera can get by panning or tilting to the setting l_j^k. If it cannot "see" an object at l_j^k, then b_{ij}^k is 0. Otherwise, b_{ij}^k is decided by the pre-defined metrics, such as the view, size and position of the object. The order of elements in $\boldsymbol{b_{ij}}$ implies the willingness of the camera to follow an object or not. We prefer to use a camera without any panning or tilting, since panning and tilting make some frames blurred and it takes time to have a sharp image. If an object is moving at a high speed, when the camera can have a sharp image after panning or tilting a large degree of angle, the object may already be out of the FOV again. However, the necessity of having this vector representation instead of by considering the current location l_j^1 only lies in the fact that in some cases, all the cameras that can currently "see" the object have a back or side view of the object while if we pan or tilt some camera, which is currently unavailable for this object, it will have the object's frontal view, which can provide us more information of interest. Or, there might be the case when a camera pans or tilts to another setting, it will gain more welfare by following another object instead of continuously following the object currently assigned to it. This vector representation helps to take into account the inclination of a camera, which, therefore, avoids the drawbacks of greedy algorithms. Finally, the overall bid price B is calculated as a function of all the intermediate bids in $\boldsymbol{b_{ij}}$, i.e.

$$B_{ij} = f(b_{ij}^1, b_{ij}^2, \dots, b_{ij}^k, \dots, b_{ij}^{24}).$$

This function is designed in the next subsection.

3. Bid submission. After evaluating the price for each object, all the related cameras send their bid prices for the object(s). As mentioned, the prices must be honest and can truly imply their willingness to follow an object.

4. Close of auction. The virtual auctioneer chooses the camera with the top bid to follow an object. Note that the highest computational load, the calculation of bid prices, is distributed to each camera node and, thus, done locally.

2.13 ***Optimality Discussion:*** Intuitively, under the assumption that the cameras are rational and honest, all the cameras report their true evaluations of the object to be tracked to the virtual auctioneer. The virtual auctioneer can, thus, obtain the maximal benefit by "selling" the item (the object to be tracked) to the camera that has the highest evaluations on the object. From the cameras' viewpoint, this transaction is optimal, since the camera which has the highest evaluation wins the right to track the object. Also, from the virtual auctioneer's standpoint, it can obtain the highest "payment" from the winner. If any agent in a system cannot increase its well-being without damaging others' well-beings, we say that it is *Pareto optimum* [12]. The fact that the cameras always reveal their true evaluation of the object to be tracked validates that the Pareto optimality of the camera grouping system is always achievable.

2.2. Metrics and Price Function Design

For the metrics used for evaluating the bids, we mainly consider the size of the person and the position of the person in the camera image, which are described as follows:

1. *The size of the tracked person,* M_{ij1}, measured by the ratio of the number of pixels inside the bounding box of the person to that of the size of the image.
2. *The position of the person in the FOV of a camera,* M_{ij2}. It is measured by the Euclidean distance that a person is away from the center of the image.

Each intermediate bid b_{ij}^k is decided by the above metrics and is calculated as $b_{ij}^k = \sum_{m=1}^{2} w_m M_{ijm}$, where w_m is the weight for different metrics. The final bid price B_{ij} is computed as

$$B_{ij} = \left(\alpha_1 (b_{ij}^1)^\lambda + \alpha_2 (b_{ij}^2)^\lambda + \cdots + \alpha_{24} (b_{ij}^{24})^\lambda\right)^{\frac{1}{\lambda}} \quad (1)$$

where $\alpha_1 + \alpha_2 + \cdots + \alpha_{24} = 1, \lambda \in (-\infty, +\infty)$.

The bid price function B_{ij} implies the utility that Camera C_j would obtain if it is assigned to follow Person P_j.

The parameter λ in equation (1) measures the degree of easiness in substitution among different dimensions in the intermediate bid vector $\boldsymbol{b_{ij}}$, i.e., when multiple setting of a camera can cover the object to be followed, to what extent we can use one of these available settings to substitute among one another in terms of the cost and benefit the camera can get. It can be proved that as λ approaches to negative infinity, B_{ij} is determined by the $\alpha_k b_{ij}^k$ with the *lowest* value. On the other hand, if λ equals 1, each dimension b_{ij}^k is a perfect substitution for any other dimension in $\boldsymbol{b_{ij}}$, i.e. each setting of the camera will give exactly the same result. Finally, as λ goes to positive infinity, the bidder's utility level is determined by $\alpha_k b_{ij}^k$ with the *highest* value.

In addition, α_k in the bid price function measures the camera's relative preference on b_{ij}^k to other b_{ij}^n $(n \neq k)$. The larger the α_k is, the larger weight is put on b_{ij}^m in the bid price function B_{ij}. In our experiments, we put the highest weight on α_1, which means that we prefer to use a camera to follow a person without any active control to avoid blurring.

The zoom control is done when a person's frontal view is detected around the centroid of an assigned camera. We zoom in that camera (if more than one are available for the frontal view, then we zoom in the one that provides a higher bid) for 2 frames and then zoom out (in case that some other person will be lost when zooming in the camera).

3. EXPERIMENTAL RESULTS

We perform the experiments in a network with 37 outdoor cameras. All the cameras are commercially available Axis 215 PTZ cameras. The map of the camera network is given in Figure 2, with the cameras used in the experiments marked in red. Some of the cameras' FOVs are overlapped while some are non-overlapped.

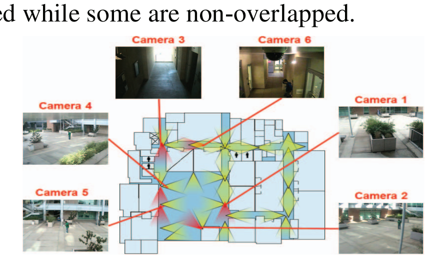


Figure 2: Map of the camera network.

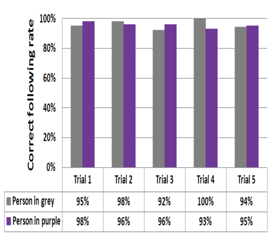
We calculate the homographies for different settings of cameras beforehand, such that we know the correspondence between each pair of cameras for any setting, based on the same ground plane. We estimate the person's location and actual height by using a camera's location above the ground plane. Thus, we estimate M_{ij1} and M_{ij2}.

We apply a particle filter tracker and use the RGB color as the feature vector. The face detection is done by applying the face detector in OpenCV in the top half of the bounding box of a person. We choose a particle filter tracker because of its robustness to occlusions. Note that the focus of this paper lies in how to form groups of cameras dynamically and integrate camera active control into this process. The parameters in the experiments are set empirically. The weights for different metrics are selected as $w_1 = 0.6$ and $w_2 = 0.4$. The elasticity of substitution parameter in the bid price function B_{ij}, $\lambda = 8$. α_k's are chosen from [0, 1] with $\sum_{k=1}^{24} \alpha_k = 1$.

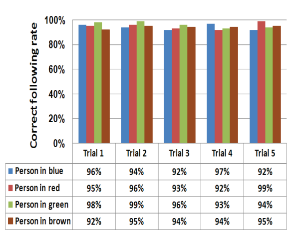
In Figure 4, we show the overall performance of the proposed approach in 2 cases: the 3 cameras 2 persons case and the 6 cameras 4 persons case. We define the correct following rate of a person as the ratio of the number of frames that a person is successfully followed when the person is visible in the network to the number of frames of the video sequence. In Figure 3, we show some typical frames in the 6 cameras and 4 persons. As stated previously, QOVs in the cameras influence the proposed camera active control results. We can observe that, when there are more than one camera available for a person, the system chose the best (potentially) available camera. For instance, in frame c, both camera 1 and camera 4 can see the person in green, but camera 6 has a better view and submit a higher bid, so it wins the auction and is selected follow the person in green. From frame c to frame d, since camera 3 is the only camera that can see the person in brown, as the person is walking forward, the camera is tilted to follow him better. In frame e, the frontal view of the person in red is detected, so, in frame f, a zooming-in operation is made to have a close-up view of the person in red. There are also some cases that a person might be lost due to the active control of a camera. For instance, in frame b, the person in red is in the FOV of camera 1, but that camera is panning and has a blurred image, which makes the person in red to get lost.

Figure 3: Experimental results in the 6 cameras 4 persons case. Figures are boxed in the same color as the person that they are assigned to follow.

Camera 1	Camera 2	Camera 3
Camera 4	Camera 5	Camera 6



(a)



(b)

Figure 4: Correct following rate for (a): the 3 cameras 2 persons case and (b) the 6 cameras 4 persons case in 5 trials.

4. CONCLUSIONS

We proposed a novel auction-based mechanism to actively control multiple cameras to follow objects in a video network. This paper introduced the auction concept into the camera network area and achieved promising results. We made the bid as a vector to take into account the cameras' willingness to follow an object or not. We show results for following various numbers of persons and active control of cameras. Experiments in real-time (15-19 fps) with real data are performed, which show the effectiveness of the proposed approach.

5. ACKNOWLEDGEMENT

This work was partially supported by NSF grants 0551741 and 0622176, 0905671 and ONR grants (DoD Instrumentation and the Aware Building). The contents of the information do not reflect the position or policy of the US Government.

REFERENCES

[1] L. Tessens *et al*. Principal view determination for camera selection in distributed smart camera networks. *ICDSC* 2008.

[2] Y. Li and B. Bhanu. Utility-based dynamic camera assignment and hand-off in a camera network. *ICDSC* 2008.

[3] R. Collins *et al*. An active camera system for acquiring multi-view video. ICIP 2002.

[4] S. K *et al*. Real-time video tracking using PTZ cameras. SPIE 6th Int. Conference on Quality Control by Artificial Vision, 2003.

[5] S. Sinha *et al*. Towards Calibrating a Pan-Tilt-Zoom Camera Network. ECCV Wkshp. Omnidirectional Vision and Camera Networks, 2004.

[6] L. He and T.R. Ioerger. Task-oriented computational economic-based distributed resource allocation mechanisms for computational grids. *ICAI*, Vol. 1, 2004, pp. 462–468.

[7] B.P. Gerkey and M.J. Mataric. Sold!: Auction methods for multirobot coordination. *IEEE T-RA*, Vol. 18, No. 5, October, 2002.

[8] M.B. Dias and A. Stentz. Opportunistic optimization for market-based multirobot control. *IROS*, 2002.

[9] Chen *et al.* Auction-based dynamic coalition for single target tracking in wireless sensor networks. *The sixth world congress on intelligent control and automation*, 2006, Vol. 1, pp. 94-98.

[10] F.Z. Qureshi and D. Terzopoulos. Distributed coalition formation in visual sensor networks: a virtual vision approach. *DCOSS, 2007*.

[11] R. G. Smith. The Contract net protocol: high-level communication and control in a distributed problem solver. *IEEE Transactions on Computers*, Vol. C-29, No. 12, December 1980.

[12] H. R. Varian. Microeconomic analysis (3rd edition). *W.W.Norton & Company*, New York, 1992.