OPTIMAL SENSOR SELECTION FOR VIDEO-BASED TARGET TRACKING IN A WIRELESS SENSOR NETWORK

Peshala V. Pahalawatta, Thrasyvoulos N. Pappas, and Aggelos K. Katsaggelos

Department of Electrical and Computer Engineering
Northwestern University, Evanston, IL 60208, USA
(pesh, pappas, aggk)@ece.northwestern.edu

ABSTRACT

The use of wireless sensor networks for target tracking is an active area of research. Imaging sensors that obtain video-rate images of a scene can have a significant impact in such networks, as they can measure vital information on the identity, position, and velocity of moving targets. Since wireless networks must operate under stringent energy constraints, it is important to identify the optimal set of imagers to be used in a tracking scenario such that the network lifetime is maximized. We formulate this problem as one of maximizing the information utility gained from a set of sensors subject to a constraint on the average energy consumption in the network. We use an unscented Kalman filter framework to solve the tracking and data fusion problem with multiple imaging sensors in a computationally efficient manner, and use a lookahead algorithm to optimize the sensor selection based on the predicted trajectory of the target. Simulation results show the effectiveness of this method of sensor selection.

1. INTRODUCTION

Video-based tracking for applications, such as surveillance and traffic analysis, is a well-studied topic in computer vision. Systems with multiple cameras can be used to detect and localize moving objects while also providing vital information on the objects, such as shape, color, and size, which can be used for identification. [1] and [2] are examples of multiple camera tracking systems proposed for wired networks of cameras. The computational cost associated with image and video processing, however, has driven recent research on tracking with wireless sensor networks towards less complex and less energy consuming sensors.

An analysis of the energy consumption and information rate of visual sensors relative to other sensors is provided in [3]. It is shown in [3] that, while visual sensors cost more in terms of energy consumption per sample than acoustic or seismic sensors, they can still be valuable partners in a wireless sensor network. Therefore, it is important to develop scalable schemes for collaboratively using multiple imaging sensors in a sensor network.

In this paper, we demonstrate a method for selecting the optimal set of imaging sensors to use for tracking a moving target such that the energy consumed in the network is within a reasonable constraint. In order to achieve this, we use the concept of *sensor utility* [4]-[6]. In [4], this idea is introduced in terms of selecting subsets of sensors at each time step such that the sum utility gain over the total time is maximized subject to a total power constraint for each sensor. However, they use a simple model for the utility function where it is assumed to monotonically increase with the number of sensors.

In [5], and [6], the tracking problem is formulated in a Bayesian framework and the information utility of a sensor is defined in terms of the posterior state distribution of the target given that a particular sensor is used. Since a measurement from the sensor is not available at the time that the decision to pick the sensor is made, an expected posterior state distribution is found for each sensor. In [7], it is shown that the expected posterior uncertainty in the target state estimate can be quantified as a conditional entropy conditioned on all the previous measurements and the new measurement. They also show that the expected posterior entropy, given a particular sensor is used, is inversely related to the mutual information between the target state and the sensor measurement. If a Gaussian state space model is assumed, then the posterior entropy does not depend on a measurement taken from the sensor, and it is directly related to the covariance of the posterior estimate.

The above method of calculating the information utility has been used in [8] and [9]. In [9], a multistep lookahead algorithm is used which maximizes the sum of the information gain over a finite horizon. This enables the system to continue to track a target through regions which may not be covered by any sensor in the network.

In the above work, the sensing is performed by one *leader node* at each time step. A significant difference in using imaging sensors is that measurements from at least two sensors at a time are required in order to continue to accurately triangulate the position of the target. Therefore, our algorithm must choose the best subset of two or more sensors from among all the sensors that could potentially be able to sense the target. Another important difference is that imaging sensors generally rely on motion segmentation in order to detect the target in the scene. Since motion segmentation algorithms require some initialization time, our algorithm must look ahead over an adequate number of steps so that it can turn on a potentially useful sensor prior to the time when the sensor is actually utilized. Here, we assume that sensors use energy for processing as well as the transmission of information.

This paper is organized as follows. In the next section, we briefly describe the tracking method based on the Unscented Kalman filter [10], [11] that was utilized in our work. We show how the sensor utility measure can be calculated as part of the tracking approach. Section 3 describes the sensor network that is envisioned, and provides the problem formulation. In section 4, we show some simulation results where we compare the proposed approach to other possible approaches for sensor selection, and then we provide our conclusions in section 5.

2. THE TRACKING METHOD AND SENSOR UTILITY

2.1. Bayesian Formulation

In a Bayesian framework, the problem of tracking a target can be expressed as the task of finding the *a posteriori* distribution of the current target state (at time k) $\mathbf{x}(k)$, denoted, $p(\mathbf{x}(k)|\mathbf{Z^k})$, where, $\mathbf{Z^k}$ denotes all the measurements taken up to time k. The above distribution can be written recursively as:

$$p(\mathbf{x}(k)|\mathbf{Z^k}) = \frac{p(\mathbf{z}(k)|\mathbf{x}(k))p(\mathbf{x}(k)|\mathbf{Z^{k-1}})}{p(\mathbf{z}(k)|\mathbf{Z^{k-1}})} \tag{1}$$

where $\mathbf{z}(k)$ is a vector of measurements taken by the subset of sensors that is active at time k. We consider the target state, $\mathbf{x}$, to consist of the x,y, and z coordinates of the target and the associated velocities, x_v, y_v and z_v in each direction. Thus, the state vector can be written as:

$$\mathbf{x} = [x, x_v, y, y_v, z, z_v]^T \tag{2}$$

The observations will be the u, v coordinates of the target as observed by the image plane of each sensor. Therefore, the observation vector for each sensor i can be written as:

$$\mathbf{z_i} = [u_i, v_i]^T \tag{3}$$

The observation function can be described using the projective matrix Pr_i, obtained by calibrating the sensor. In previous work, we have demonstrated a method for self-calibrating a set of imaging sensors in a network [12]. Therefore, we can represent the relationship between the observations and the target state as follows:

$$\begin{bmatrix} u_i S \\ v_i S \\ S \end{bmatrix} = Pr_i \begin{bmatrix} x \\ y \\ z \\ 1 \end{bmatrix} \tag{4}$$

From eq. (4), it is clear that the observation (u, v) at each sensor is a non-linear function of the target state since the scaling parameter, S, depends on the target state.

2.2. The Unscented Kalman Filter

Due to the non-linearity of the observation model, a standard linear Kalman filter approach cannot be used to solve our tracking problem. While a possible approach would be to linearize the observation model and develop an extended Kalman filter framework, an alternative method that is simple to implement and produces better estimates of the first and second order statistics of the target state is presented in [10], and [11]. In this paper, we will not delve into extensive details of this approach, which is termed the *Unscented Kalman Filter* (UKF). The basic idea of the UKF can be stated as that of generating a set of samples of a known probability distribution (i.e., a set of samples that will have the same mean and covariance), and then of transforming the sample set according to the dynamic and observation models of the system. The transformation, which is termed an *unscented transformation*, satisfies certain properties that ensure the resulting statistics will be consistent and the resulting estimates will be unbiased. Since a Gaussian distribution is completely described by its mean and covariance, the number of samples required in order to represent the distribution is finite and small.

We assume that the dynamic model for the target state $\mathbf{x}$ and the observation model for the sensors can be specified respectively as:

$$\mathbf{x}(k+1) = f(\mathbf{x}(k), \mathbf{v}(k)), \tag{5}$$

and,

$$\mathbf{z}_{\mathcal{C}}(k+1) = h_{\mathcal{C}}(\mathbf{x}(k+1), \mathbf{w}(k+1)) \tag{6}$$

where noise vectors $\mathbf{v}(k)$ and $\mathbf{w}(k)$ are zero mean, uncorrelated Gaussian processes. The observation function, $h_{\mathcal{C}}$, is a non-linear function defined by (4) where $\mathcal{C}$ denotes the subset of sensors whose measurements are taken.

Now, given the current state $\mathbf{x}(k)$, current state covariance $\mathbf{P}(k)$, and a set of sensors $\mathcal{C}$, the UKF can output $\mathbf{P}_{\mathcal{C}}(k+1)$, which is the covariance of the posterior distribution of $\mathbf{x}(k+1)$ given that the UKF is updated with measurements from the sensor subset $\mathcal{C}$. As with the standard linear Kalman filter, the actual sensor measurements are not required in the calculation of $\mathbf{P}_{\mathcal{C}}(k+1)$. Once the actual sensor measurements are obtained, the UKF can be used to estimate the mean $\hat{\mathbf{x}}(k+1)$.

2.3. Sensor Utility Measure

A larger covariance in the posterior distribution of $\mathbf{x}(k+1)$ implies a greater uncertainty in the estimate $\hat{\mathbf{x}}(k+1)$ that would be given by the tracker. Therefore we can use the trace of the covariance matrix $\mathbf{P}_{\mathcal{C}}(k+1)$ as a measure of the uncertainty in the estimate $\hat{\mathbf{x}}(k+1)$ given that observations are to be taken from the sensor subset $\mathcal{C}$. Thus, we can quantify the information utility, $\psi_{\mathcal{C}}(k+1)$, of the sensor set $\mathcal{C}$.

$$\psi_{\mathcal{C}}(k+1) = -trace[\mathbf{P}_{\mathcal{C}}(k+1)] \tag{7}$$

3. PROBLEM FORMULATION

3.1. The Sensor Network

We consider a network of N arbitrarily placed and arbitrarily oriented imaging sensors. Each sensor is assumed to have a limited field of view, defined by the intrinsic parameters and orientation of that sensor. The sensors are assumed to be calibrated prior to the tracking task.

The network also includes low-power sensors (eg., passive infra-red) that can act as triggers to initially power on the imaging sensors when a target is approaching the perimeter of the sensor field. Each low-power sensor is assumed to be capable of triggering its neighboring imaging sensors. We also assume that the target state estimates are maintained by a set of processing nodes that are located within each cluster of sensors. Each processing node can send and receive transmissions to and from all its neighboring imaging sensors as well as its neighboring processing nodes.

For the purposes of this paper, we assume that the sensors can transmit at only one power level with which they are able to communicate with all the sensors within their neighborhood, and also that the sensors are homogeneous in terms of their processing and transmission energy consumption.

Since a sensor cannot transmit any useful information while it is still in its initialization stage, we assume that each imaging sensor i can be in one of three possible states, π_i, at time k. They are:

- $\pi_i[k] = 0$: Sensor is off (inactive)
- $\pi_i[k] = 1$: Sensor is on and processing but not transmitting (initializing)

- $\pi_i[k] = 2$: Sensor is on and transmitting

We can assume that each sensor needs an initialization time of T time steps. Therefore, the energy used by each sensor i at time k is a function of the sensor state $\pi_i[k]$, which in turn is constrained by the sensor state over T previous time steps.

3.2. Problem Definition

We define the optimal sensor selection problem as that of selecting the states of the set of sensors throughout the tracking period such that the sum information utility provided by the active sensors will be maximized subject to a constraint on the average energy that can be used by the network. The average energy constraint allows the network to limit the number of sensors that will be active and transmitting at any given time while still allowing for some flexibility in order to maintain the tracking accuracy.

$$\begin{aligned} \min_{\pi_{\mathcal{N}}} \quad & \sum_{k=1}^{K} -U(\pi[k], k) \\ s.t. \quad & \\ & \sum_{k=1}^{K} \sum_{i \in N} E_i(\pi_i[k]) \leq K \cdot E_{ave} \end{aligned} \tag{8}$$

where, $U(\pi[k], k)$ is the information utility provided by the network state $\pi[k]$ at time k. K is the total time the target remains within the field of view of the network, and E_{ave} is the average energy constraint. Note that if $\mathcal{C}$ is the set of sensors i such that $\pi_i[k] = 2$ (i.e. on and transmitting at time k), then $U(\pi[k], k)$ would be equal to $\psi_{\mathcal{C}}(k)$ as provided by (7).

Since the network can potentially consist of a large number of sensors, the above problem is difficult to solve in a centralized manner. The total time the target remains within the network can also be long, and the information utility of the sensors can only be tracked over a finite time based on the accuracy of the target dynamic model. Therefore, it is important to solve the above problem in a distributed manner over finite lengths of time. Our proposed approach solves the problem in a distributed manner by using the processing nodes in the network. We make the problem tractable by optimizing over a finite window, $W = T + 1$, in time, where T is the initialization time for each sensor. Therefore, in the proposed method, this optimization is performed at each time step by the processing node that is closest to the target based on the previously estimated target state. The processing node calculates the current target state, and hands over the processing to the next node that will be closest to the target. Each processing node has the capability of receiving transmissions only from the sensors in the network that are within its range. Now, the problem can be stated as follows:

At each time step, k, pick the processing node closest (based on the previous state estimate) to the target. Let the sensors within range of the processing node be in the set $\mathcal{R}$. Then, we define the objective to be minimized by the processing node at each time step as:

$$\begin{aligned} \min_{\pi_{\mathcal{R}}[k]} \quad & -U(\pi_{\mathcal{R}}[k], k) - \sum_{j=k+1}^{k+W-1} \hat{U}(\pi_{\mathcal{R}}[j], j) \\ s.t. \quad & \\ & \sum_{j=k-W+1}^{k} \sum_{i \in N} E_i(\pi_i[j]) \leq W \cdot E_{ave} \\ & \sum_{i \in N} E_i(\pi_i[j]) \leq E_{pk} \end{aligned} \tag{9}$$

Here E_{pk} is a peak energy constraint that essentially limits the maximum number of sensors that can be used at any time. The term $\hat{U}(\pi_{\mathcal{R}}[j], j)$ denotes the estimated information utility for future time steps, which will depend on future choices of the network state π. The above minimization problem is solved using a tree-pruning algorithm. Since the computational complexity of the algorithm increases exponentially with W, and since our confidence in the expected utility measure decreases with time, we also limit the number of nodes in the tree that are preserved at each time step to be within a maximum bound M. Essentially, at each time step, the nodes of the tree are sorted according to their sum information utility values up to that time step, and only the top M nodes are preserved.

Once the correct sensor states are chosen, the processing node will transmit a command to each relevant sensor to configure its state. The processing node will then update its estimate of the target state based on measurements received from the selected sensors. Once the estimate is updated, the processing node will either hand-off the processing task to a new node that is closer to the target than itself, or, if it is still the closest node, it will continue the tracking and sensor selection process.

4. SIMULATIONS

To demonstrate the feasibility of the proposed sensor selection approach, we simulated a network with imaging sensors laid out in a 50m x 50m area, divided into 10m x 10m blocks. Each block contained 2 randomly placed and oriented imaging sensors. The center of each block contained a processing node. We assumed that each processing node could communicate with any sensor or processing node within its own block, or its 8 neighboring blocks. A processing node becomes active and begins to optimally select sensor states only when the estimated target position is within its block.

For our experiments, the sensor frame rate was assumed to be 30fps. The resolution of the sensors was set at 1280x1024 pixels, and the focal lengths were set at 5.4mm. The target was assumed to be about the size of a medium-sized car with dimensions of about 4.8m x 1.8m x 1.5m (length, width, height). The observation from each sensor was the perceived centroid of the target for that sensor. The target speed was set between 30-60m/sec, and the tracking period was 50sec.

Figure 1 shows an example setup of the sensor network. We also show the estimated target trajectory, vs the ground truth using our tracking and sensor selection method for a specific realization of sensor placements and target trajectory. For this simulation, we set the initialization time to 2 time steps.

To demonstrate the advantages of the proposed approach, we first compared the proposed method to a random selection approach. In the random approach, the transmitting and initializing sensors were chosen randomly from the set of sensors for which the target would be within the field of view based on its estimated location. The initialization time was set to one time step. The comparison was based on results averaged over 100 complete independent runs with each run having a random sensor placement and target trajectory. A complete run was considered to be a run in which neither method lost track of the target. A target track was assumed to be lost if the target was not seen by any sensor for at least 5 time steps. The results for each method are shown in fig. 2. The actual average energy used by the random sensor selection method tends to be lower than the optimal method since it does not initialize sensors that could potentially be of use in the next time step. The rmse values are averaged over only the realizations in

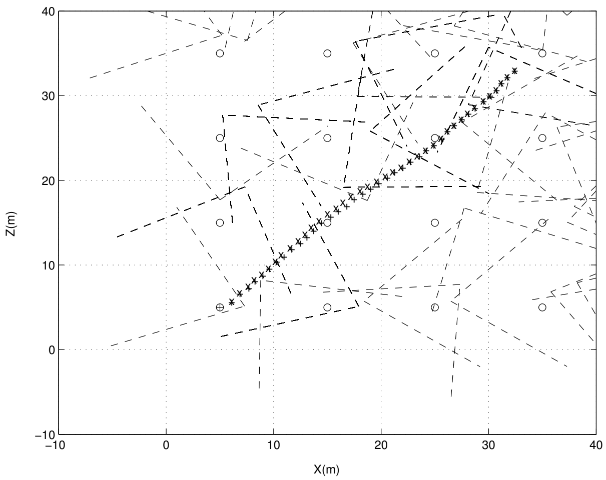

Fig. 1. The sensor network: the dashed lines indicate the field of view of each sensor, and the circles indicate the positions of the processing nodes. The '+' indicates the ground truth for the target trajectory, and the 'x' indicates the estimated trajectory.

	Optimal Selection	Random Sensor Selection
Ave. E	4.93	2.41
rmse	0.40	0.52
% lost tracks	6.3	58.0

Fig. 2. Comparison between proposed sensor selection approach based on predicted utility gain over three steps, and random sensor selection approach. The average energy threshold was set at 5.5, and initialization time was set at 1 time step.

which both approaches were able to track the target throughout its entire trajectory.

We also compared our approach with a different approach that uses the distance from the estimated target position to the position of each sensor as a measure of the sensor utility. In this method, the sensors closest to the target are assumed to have the greatest utility in observing its location. The two approaches were compared under similar conditions with the the same constraints on energy, and an initialization time of two time steps. The results for the two approaches are shown in fig. 3. Again, the results are averaged over 100 complete independent runs with each run having a different random sensor placement and target trajectory. The information utility based approach performs better in terms of the tracking accuracy when compared with the other approach.

	Optimal Selection	Closest Sensor Selection
Ave. E	4.87	4.41
rmse	0.44	0.52
% lost tracks	10.3	11.2

Fig. 3. Comparison between information based utility approach and distance based utility approach. The average energy threshold was set at 5.5, and initialization time was set at 2 time steps.

5. CONCLUSIONS

In this paper, we have developed a feasible approach for target tracking in a wireless sensor network using imaging sensors. We have developed a method that takes into account the initialization time required by imaging sensors in order to perform motion segmentation for target detection and tracking. We have shown the gain in terms of tracking accuracy that can be acheived by selecting imaging sensors based on an information utility criterion subject to the energy constraint. We have also demonstrated that the sensor utility can be obtained in a computationally efficient manner using an unscented Kalman filter.

6. ACKNOWLEDGEMENTS

This project has been funded in part by the Defense Advanced Research Projects Agency (DARPA) through the Sensor Information Technology Program (SensIT).

7. REFERENCES

[1] Kogut, G.T., Trivedi, M.A.: Maintaining the identity of multiple vehicles as they travel through a video network. Proc. of Intelligent Transportation Systems (2001) 756–761

[2] Khan, S, Shah, M.: Consistent labeling of tracked objects in multiple cameras with overlapping fields of view. IEEE Trans. on Pattern Analysis and Machine Intelligence **25** (10) (2003) 1355–1360

[3] Boettcher, P.W., Shaw, G.A.: Energy-constrained collaborative processing for target detection, tracking and geolocation. Proc. of IPSN (2003) 254–268

[4] Byers, J., Nasser, G.: Utility-based decision making in wireless sensor networks. Proc. of IEEE MobiHOC (2000)

[5] Zhao, F., Shin, J., Reich, J.:Information-driven dynamic sensor collaboration for target tracking. IEEE Signal Processing Magazine **19** (2) (2002)

[6] Chu, M., Haussecker, H., Zhao, F.: Scalable information-driven sensor querying and routing for ad hoc heterogeneous sensor networks. Int'l J. of High-Performance Computing Applications **16** (3) (2002) 90–110

[7] Ertin, E., Fisher, J.W., Potter, L.C.: Maximum mutual information principle for dynamic sensor query problems. Proc. of IPSN (2003) 405–416

[8] Liu, J., Reich, J., Zhao, F.: Collaborative in-network processing for target tracking. Eurasip J. on Applied Signal Processing **4** (2003) 378–391

[9] Liu, J., Petrovic, D., Zhao, F.: Multi-step information-directed sensor querying in distributed sensor networks. Proc. of ICASSP **5** (2003) 145–148

[10] Julier, S.J., Uhlmann, J.K.: A new extension of the Kalman filter to nonlinear systems. Int. Symp. Aerospace/Defense Sensing, Simul. and Controls (1997)

[11] Wan, E.A., Van der Merwe., R.: The Unscented Kalman Filter. Kalman Filtering and Neural Networks, ed. Simon Haykin, Wiley (2001)

[12] Pahalawatta, P.V., Depalov, D., Pappas, T.N., Katsaggelos, A.K.: Detection, classification, and collaborative tracking of multiple targets using video sensors. Proc. of IPSN (2003) 529–544