

RESEARCH

Open Access



Multi-camera object tracking using surprisal observations in visual sensor networks

Venkata Pathuri Bhuvana^{1,2*}, Melanie Schranz¹, Carlo S. Regazzoni², Bernhard Rinner¹, Andrea M. Tonello¹ and Mario Huemer³

Abstract

In this work, we propose a multi-camera object tracking method with surprisal observations based on the cubature information filter in visual sensor networks. In multi-camera object tracking approaches, multiple cameras observe an object and exchange the object's local information with each other to compute the global state of the object. The information exchange among the cameras suffers from certain bandwidth and energy constraints. Thus, allowing only a desired number of cameras with the most informative observations to participate in the information exchange is an efficient way to meet the stringent requirements of bandwidth and energy. In this paper, the concept of surprisal is used to calculate the amount of information associated with the observations of each camera. Furthermore, a surprisal selection mechanism is proposed to facilitate the cameras to take independent decision on whether their observations are informative or not. If the observations are informative, the cameras calculate the local information vector and matrix based on the cubature information filter and transmit them to the fusion center. These cameras are called as surprisal cameras. The fusion center computes the global state of the object by fusing the local information from the surprisal cameras. Moreover, the proposed scheme also ensures that on average, only a desired number of cameras participate in the information exchange. The proposed method shows a significant improvement in tracking accuracy over the multi-camera object tracking with randomly selected or fixed cameras for the same number of average transmissions to the fusion center.

Keywords: Kalman filters, Information filters, State estimation, Information entropy

1 Introduction

Object tracking is an extensively studied topic in visual sensor networks (VSN). A VSN is a network composed of smart cameras; they capture, process, and analyze the image data locally and exchange extracted information with each other [1]. The main applications of a VSN are indoor and/or outdoor surveillance, e.g., airports, massive waiting rooms, forests, deserts, inaccessible locations, and natural environments [2]. In general, the typical task of a VSN is to detect and track specific objects. The objects are usually described by a state that includes various

characteristics of the objects such as position, velocity, appearance, behavior, shape, and color. These states can be used to detect and track the objects. Recursive state estimation algorithms are predominantly used to track objects in a VSN [3].

In [4–11], the authors presented several Kalman filter (KF)-based object tracking methods. Extended Kalman filter (EKF)-based object tracking method is proposed in [12]. The unscented Kalman filter (UKF) is applied for visual contour tracking in [13] and object tracking in [14]. In terms of object tracking in a VSN, the cubature Kalman filter (CKF) is primarily applied in our previous work [15]. In [16–24], the authors presented particle filter (PF)-based object tracking. The object tracking methods based on these conventional Bayesian filters have a varying degree of complexity and accuracy.

*Correspondence: venkata.pathuri@aau.at

¹Institute of Networked and Embedded Systems, Alpen-Adria-Universität, Klagenfurt, Austria

²Department of Marine Engineering, Electrical, Electronics, and Telecommunications, University of Genova, Genova, Italy

Full list of author information is available at the end of the article

In general, the performance of the tracking algorithms suffers from different adverse effects such as distance or orientation of the camera, and occlusions. However, a VSN with overlapping field of views (FOVs) is capable of providing multiple observations of the same object simultaneously. The authors in [25] presented a distributed and collaborative sensing mechanism to improve the observability of the objects by dynamically changing the camera's pan, tilt, and zoom. Other examples of distributed object tracking methods are presented in [26] and [27].

Recently, information filters have emerged as suitable methods for multi-sensor state estimation [28]. In information filtering, the information vector and matrix are computed and propagated over time instead of the state vector and its error covariance. The information matrix is the inverse of the state error covariance matrix. The information vector is the product of the information matrix and state vector. The information filters have an inherent information fusion mechanism which makes them more suitable for multi-camera object tracking. A more detailed description of information filters is given in Section 3. The authors in [29] and [30] presented information weighted consensus-based distributed object tracking with an underlying KF or a distributed maximum likelihood estimation. In our work [31], we have presented a robust cubature information filter (CIF)-based distributed object tracking in VSNs. However, the limited processing, communication, and energy capabilities of the cameras in a VSN present a major challenge.

Nowadays, VSNs tend to evolve into large-scale networks with limited bandwidth and energy reservoirs. This allows a large number of cameras to observe a single object. In spite of the improved tracking accuracy, the information exchange of the large number of observations among the cameras increases the communication overhead and energy consumption. Hence, allowing only a desired number of cameras to participate in the information exchange is a way to meet the stringent requirements of bandwidth and energy.

Estimating an object's state with a selected set of cameras is a well-investigated topic. Several camera selection mechanisms have been proposed in literature to minimize and/or maximize different metrics such as estimation accuracy, monitoring area, number of transmissions, and amount of data transfer. In [32], the authors presented an object tracking method based on fuzzy automaton in handing over to expand the monitoring area. This method selects a single best camera to control and track the objects by comparing its rank with the neighboring cameras. This method fails to select multiple cameras, and cameras have to communicate with each other to select the best camera. In [33], the authors presented an efficient camera-tasking approach to minimize the visual hull

area (maximal area that could be occupied by objects) for a given number of objects and cameras. They also presented several methods to select a subset of cameras based on the positions of the objects and cameras to minimize the visual hull area. If the objects are recognized in the vicinity of a certain location, then a subset of cameras that is best suited to observe this location performs the tracking. This method is capable of selecting multiple cameras but not the desired number of cameras on average. In [34], the authors presented a framework for dynamically selecting a subset of cameras to track people in a VSN with limited network resources to achieve the best possible tracking performance. However, the camera selection decision is made at the FC based on training data and the selection is broadcast to the cameras in the VSN. Hence, this selection process does not depend on the true observations.

The observations received by the cameras in the VSN are typically realizations of a random variable. Hence, they contain a varying degree of information about the state of the object. They can be broadly classified into informative and uninformative observations. The non-informative observations do not contribute significantly to the tracking accuracy. Hence, a camera selection strategy that allows only a desired number of cameras with most informative observations to participate in the information exchange and discards the cameras with non-informative observations is an efficient way to meet the requirements of bandwidth and energy.

In [35], the authors presented an entropy-based algorithm that dynamically selects multiple cameras to reduce transmission errors and subsequently communication bandwidth. In this work, the cameras in the VSN use the extended information filter (EIF) as the local filter and calculate the expected information gain (EIG) in the form of a logarithmic ratio of the expected and posterior information matrices. If the information gain is greater than the cost of transmissions, then the cameras participate in the information fusion. The calculated EIG in this method does not depend on the measurements directly, and the cluster head has to run an optimization step to select the best possible cameras at each step. Moreover, this method is not capable of selecting only a desired number of cameras on average. In [36], a camera set is selected based on an individual image quality metric (IQM) for spherical objects. The cameras that detect the spherical target are ranked in ascending order based on their value of the local IQM, and the required number of cameras with highest IQM are chosen. This approach is limited to spherical objects. However, it can be easily extended to non-spherical objects. The major disadvantage of this method is either all the cameras in the VSN or the FC should know IQM of all the other cameras in the VSN. Hence, this method does not ensure

cameras to take independent decisions thus restricting the scalability.

In our work, a multi-camera object tracking method based on the CIF is proposed in which the cameras can take independent decisions on whether or not to participate in information exchange. Furthermore, the proposed method also ensures that on average, only a desired number of cameras participate in the information exchange to meet bandwidth requirements. We model the state of an object utilizing a dynamic state representation that includes its position and velocity on the ground plane. Further, we consider a VSN with overlapping FOVs; thus, multiple cameras can observe an object simultaneously. Each camera in the VSN has a local CIF on board. Hence, they can calculate the local information metrics (information contribution vector and matrix) based on their observations. The cameras that can observe a specific object form a cluster (observation cluster) with an elected fusion center (FC). In this paper, we consider the concept of surprisal [37] to evaluate the amount of information in the observations received by the cameras in the VSN. The surprisal of the measurement residual indicates the amount of new information received from the corresponding observation. The observations of a camera are informative only if the corresponding surprisal of the measurement residual is greater than a threshold. The threshold is calculated as a function of the ratio of the number of desirable cameras and the total number of cameras in the observation cluster. This ensures that on average, only the desired number of cameras are selected as the cameras with informative observations (surprisal cameras). The surprisal cameras calculate the local information metrics based on the CIF and transmit them to the FC. Then, the FC fuses the surprisal local information metrics

to achieve the global state by using the inherent fusion mechanism of the CIF. The proposed selection mechanism only requires the knowledge of the total number of cameras in the observation cluster and the desired number of cameras. Further, we compare the proposed multi-camera object tracking method with surprisal cameras with multi-camera object tracking with random and fixed cameras using simulated and experimental data.

The paper is organized as follows: Section 2 describes the considered VSN with motion and observation models. Section 3 presents theoretical concepts of information filtering. Section 4 describes the camera selection based on the surprisal of the measurement residual and the calculation of the surprisal threshold. Section 5 explains the proposed CIF-based multi-camera object tracking with surprisal cameras. Section 6 evaluates the proposed method based on simulation and experimental data. Finally, Section 7 presents the conclusions.

2 System model

In this work, we consider a VSN consisting of a fixed set of calibrated smart cameras c_i , where $i \in \{1, 2, \dots, M\}$, with overlapping FOVs as illustrated in Fig. 1. The task of the cameras in the VSN is to monitor the given environment and to identify and track an object. As these cameras are calibrated, there exists a homography to calculate the object's position on the ground plane. The cameras c_i that can observe the object at time k form the observation cluster C_k . The state of the object comprises its position (x_k, y_k) and the velocity (\dot{x}_k, \dot{y}_k) on the ground plane. Thus, the state at time k is described as $\mathbf{x}_k = [x_k \ y_k \ \dot{x}_k \ \dot{y}_k]^T$. The motion model of the object at camera c_i at time k is given as

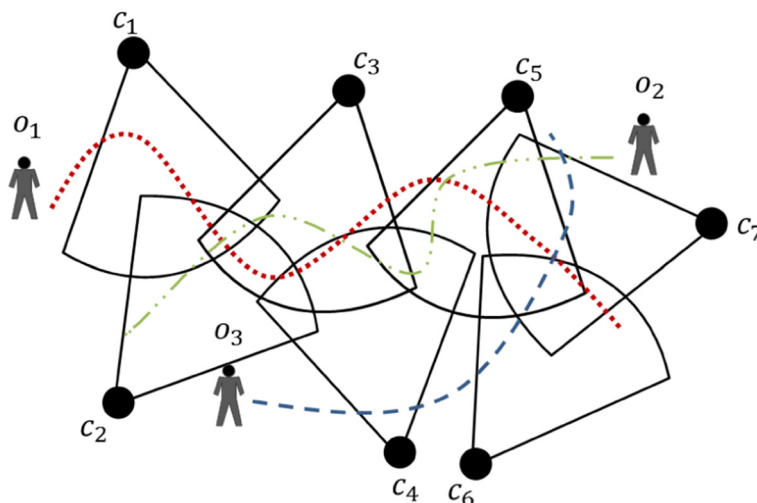


Fig. 1 Visual sensor network

$$\begin{aligned} \mathbf{x}_k &= \mathbf{f}_{i,k}(\mathbf{x}_{k-1}, \mathbf{w}_{i,k}) \\ &= \begin{bmatrix} x_{k-1} + \delta \dot{x}_{k-1} + \frac{\delta^2}{2} \ddot{x}_{i,k} \\ y_{k-1} + \delta \dot{y}_{k-1} + \frac{\delta^2}{2} \ddot{y}_{i,k} \\ \dot{x}_{k-1} + \delta \ddot{x}_{i,k} \\ \dot{y}_{k-1} + \delta \ddot{y}_{i,k} \end{bmatrix}, \end{aligned} \quad (1)$$

where \ddot{x} and \ddot{y} represent the acceleration of the object in x and y directions that are modeled by the independent and identically distributed (IID) white Gaussian noise vector $\mathbf{w}_{i,k} = [\ddot{x}_{i,k} \ \ddot{y}_{i,k}]^T$ with covariance $\mathbf{Q}_{i,k} = \text{diag}(qx_i, qy_i)$. δ is time interval between two observations. The state transition model (1) can be further written as

$$\mathbf{x}_k = \begin{bmatrix} 1 & 0 & \delta & 0 \\ 0 & 1 & 0 & \delta \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \mathbf{x}_{k-1} + \mathbf{w}_{i,k}^s, \quad (2)$$

where $\mathbf{w}_{i,k}^s$ is IID white Gaussian noise vector with covariance

$$\mathbf{Q}_{i,k}^s = \begin{bmatrix} \frac{qx_i\delta^4}{4} & 0 & \frac{qx_i\delta^3}{2} & 0 \\ 0 & \frac{qy_i\delta^4}{4} & 0 & \frac{qy_i\delta^3}{2} \\ \frac{qx_i\delta^3}{3} & 0 & qx_i\delta^2 & 0 \\ 0 & \frac{qy_i\delta^3}{3} & 0 & qy_i\delta^2 \end{bmatrix}. \quad (3)$$

The state of the object is estimated from observations taken at each time step k . The observation model of the object at camera c_i and time k is given as

$$\mathbf{z}_{i,k} = \mathbf{h}_{i,k}(\mathbf{x}_k) + \mathbf{v}_{i,k}, \quad (4)$$

where $\mathbf{v}_{i,k}$ is an IID measurement noise vector with covariance $\mathbf{R}_{i,k}$. The measurement function $\mathbf{h}_{i,k}$ is the non-linear homography function which converts the object's coordinates from the ground to the image plane. The considered motion model (1) and measurement model (4) are adapted from [27].

3 Information filtering

The information filter is an alternative version of the Bayesian state estimation methods. In information filtering, the information vector and the information matrix are computed and propagated instead of the estimated state vector and the error covariance. The estimated global information matrix $\mathbf{Y}_{k-1|k-1}$ and information vector $\hat{\mathbf{y}}_{k-1|k-1}$ at time $k-1$ are given as

$$\mathbf{Y}_{k-1|k-1} = \mathbf{P}_{k-1|k-1}^{-1}, \quad (5)$$

$$\hat{\mathbf{y}}_{k-1|k-1} = \mathbf{Y}_{k-1|k-1} \hat{\mathbf{x}}_{k-1|k-1}, \quad (6)$$

where $\hat{\mathbf{x}}_{k-1|k-1}$ and $\mathbf{P}_{k-1|k-1}$ are the estimated global state vector and error covariance matrix at time $k-1$. At time k and camera c_i , the information filter has two steps: time and measurement update.

3.1 Time update

The information form of the predicted state and the corresponding information matrix are computed as

$$\mathbf{Y}_{i,k|k-1} = \mathbf{P}_{i,k|k-1}^{-1}, \quad (7)$$

$$\hat{\mathbf{y}}_{i,k|k-1} = \mathbf{Y}_{i,k|k-1} \hat{\mathbf{x}}_{i,k|k-1}, \quad (8)$$

where $\hat{\mathbf{x}}_{i,k|k-1}$ and $\mathbf{P}_{i,k|k-1}$ are the predicted state vector and the error covariance matrix, respectively.

3.2 Measurement update

Upon receiving the measurement $\mathbf{z}_{i,k}$, the information contribution matrix $\mathbf{I}_{i,k}$ and information contribution vector $\mathbf{i}_{i,k}$ are computed as

$$\mathbf{I}_{i,k} = \mathbf{Y}_{i,k|k-1} \mathbf{P}_{\mathbf{z},i,k} \mathbf{R}_{i,k}^{-1} \mathbf{P}_{\mathbf{z},i,k}^T \mathbf{Y}_{i,k|k-1}^T, \quad (9)$$

$$\mathbf{i}_{i,k} = \mathbf{Y}_{i,k|k-1} \mathbf{P}_{\mathbf{z},i,k} \mathbf{R}_{i,k}^{-1} \left(\mathbf{e}_{i,k} + \mathbf{P}_{\mathbf{z},i,k}^T \hat{\mathbf{y}}_{i,k|k-1} \right), \quad (10)$$

where $\mathbf{P}_{\mathbf{z},i,k}$, $\mathbf{R}_{i,k}$, and $\mathbf{e}_{i,k}$ are the cross-covariance of the state and measurement vector, the measurement noise variance, and the measurement residual, respectively. The measurement residual is defined as

$$\mathbf{e}_{i,k} = \mathbf{z}_{i,k} - \hat{\mathbf{z}}_{i,k|k-1}, \quad (11)$$

where $\hat{\mathbf{z}}_{i,k|k-1}$ is the predicted measurement. In this work, the CIF is used at the cameras to track the objects locally. We refer to Appendices 1 and 2 and [38] for the CIF algorithm.

3.3 Information fusion

In multi-camera networks, multiple cameras have an overlapping FOV and thus can observe an object simultaneously. Hence, each camera c_i where $i \in C_k$ that observes the object computes its own information contribution vector $\mathbf{i}_{i,k}$ and information contribution matrix $\mathbf{I}_{i,k}$ as shown in (9) and (10), respectively. Let us consider that each camera sends their local information metrics to an elected FC, then the global information equivalents of the estimated state and error covariances at the FC c_o , where $o \in C_k$ are calculated as

$$\mathbf{Y}_{k|k} = \mathbf{Y}_{o,k|k-1} + \sum_{i=1}^{|C_k|} \mathbf{I}_{i,k}, \quad (12)$$

$$\hat{\mathbf{y}}_{k|k} = \hat{\mathbf{y}}_{o,k|k-1} + \sum_{i=1}^{|C_k|} \mathbf{i}_{i,k}, \quad (13)$$

where $\hat{\mathbf{y}}_{o,k|k-1}$ and $\mathbf{Y}_{o,k|k-1}$ are the predicted information vector and matrix at the FC, respectively.

4 Surprisal camera selection

The VSNs usually have limited bandwidth and energy reservoirs. Therefore, it might be necessary that only a desired number of cameras (subset) transmit their local information to the FC. On the other hand, this can lead to decreased tracking accuracy. A better tracking accuracy can be achieved by selecting the cameras based on the information associated with their observations. This strategy improves the accuracy of the global state estimation under the given bandwidth and energy constraints. The information content associated with the observations can be calculated by applying the concept of self-information or surprisal.

4.1 Surprisal

The surprisal H is a measure of the information associated with the outcome x of a random variable. It is calculated as

$$H = -\log(\Pr(x)), \quad (14)$$

where $\Pr(x)$ is the probability of the outcome x and the base of the logarithm can be considered as 2, 10, or e . In this paper, the surprisal is calculated with the natural logarithm (base e) for the sake of mathematical simplification. The surprisal of the outcome of a random variable depends only on the probability of the corresponding outcome $\Pr(x)$. A highly probable outcome of a random variable is less surprising and vice versa.

4.2 Surprisal of measurement residual

In multi-camera object tracking, the local observations $\mathbf{z}_{i,k}$ of each camera c_i are random variables because of the additive Gaussian noise and the random initial state. Hence, they contain a varying degree of information about the state of the object. Within the framework of information filtering, the measurement residual $\mathbf{e}_{i,k}$ at camera c_i and time k is the disagreement between the predicted observation and the actual observation (see (11)). Hence, the surprisal of the measurement residual $\mathbf{e}_{i,k}$ gives the additional information associated with the received observations that is not available in the predicted observations through the predicted state. The surprisal of the measurement residual $\mathbf{e}_{i,k}$ at camera c_i and time k can be computed as¹

$$H_{i,k} = -\log_e(p(\mathbf{e}_{i,k})). \quad (15)$$

Under the assumptions of IID additive Gaussian observation noise, the measurement residual becomes approximately a Gaussian distributed variable with zero mean and the covariance $\mathbf{P}_{\mathbf{z}\mathbf{z},i,k}$, called the innovation covariance

$$\mathbf{e}_{i,k} \sim \mathcal{N}(0, \mathbf{P}_{\mathbf{z}\mathbf{z},i,k}). \quad (16)$$

By substituting (16) in (15), the surprisal of the measurement residual $\mathbf{e}_{i,k}$ becomes

$$\begin{aligned} H_{i,k} &= -\log_e \left(\frac{\exp\left(-\frac{1}{2} \mathbf{e}_{i,k}^T \mathbf{P}_{\mathbf{z}\mathbf{z},i,k}^{-1} \mathbf{e}_{i,k}\right)}{(2\pi)^{\frac{n_z}{2}} \det^{\frac{1}{2}}(\mathbf{P}_{\mathbf{z}\mathbf{z},i,k})} \right) \\ &= \alpha_{i,k} + \frac{1}{2} \mathbf{e}_{i,k}^T \mathbf{P}_{\mathbf{z}\mathbf{z},i,k}^{-1} \mathbf{e}_{i,k}, \end{aligned} \quad (17)$$

where $\alpha_{i,k}$ is

$$\alpha_{i,k} = \frac{n_z}{2} \log_e(2\pi) + \frac{1}{2} \log_e(\det(\mathbf{P}_{\mathbf{z}\mathbf{z},i,k})), \quad (18)$$

and n_z is the length of the observation vector of camera c_i at time k . The observations of the camera c_i at time k are informative enough if the surprisal of the corresponding measurement residual $H_{i,k}$ is greater than a threshold

$$H_{i,k} \begin{cases} \geq \chi_k & \text{informative} \\ < \chi_k & \text{non-informative.} \end{cases} \quad (19)$$

The cameras with enough informative measurements are called surprisal cameras. The threshold χ_k has to be defined based on the bandwidth and energy constraints in such a way that at each time k , on average, only a given number of cameras are selected as surprisal cameras.

4.3 Surprisal threshold

Let $\mathbf{s}_k = (s_{1,k}, s_{2,k}, \dots, s_{|C_k|,k})$ be the indication vector at time k , where $|C_k|$ is the number of cameras in the observation cluster. Each element $s_{i,k}$ in the indication vector is either 1 or 0

$$s_{i,k} = \begin{cases} 1 & \text{surprisal camera} \\ 0 & \text{non-surprisal camera.} \end{cases} \quad (20)$$

From (17), (19), and (20), the average number of times a camera c_i becomes a surprisal camera is given as

$$\begin{aligned} \mathbb{E}[s_{i,k}] &= \Pr(s_{i,k} = 1) \\ &= \Pr(H_{i,k} \geq \chi_k) \\ &= \Pr\left(\alpha_{i,k} + \frac{1}{2} \mathbf{e}_{i,k}^T \mathbf{P}_{\mathbf{z}\mathbf{z},i,k}^{-1} \mathbf{e}_{i,k} \geq \chi_k\right) \\ &= \Pr\left(\mathbf{e}_{i,k}^T \mathbf{P}_{\mathbf{z}\mathbf{z},i,k}^{-1} \mathbf{e}_{i,k} \geq 2(-\alpha_{i,k} + \chi_k)\right) \\ &= \Pr\left(\mathbf{e}_{i,k}^T \mathbf{P}_{\mathbf{z}\mathbf{z},i,k}^{-1} \mathbf{e}_{i,k} \geq \beta_k\right), \end{aligned} \quad (21)$$

where $\beta_k = 2(-\alpha_{i,k} + \chi_k)$. Since $\mathbf{e}_{i,k} \sim \mathcal{N}(0, \mathbf{P}_{\mathbf{z}\mathbf{z},i,k})$,

$$\mathbf{e}_{i,k}^T \mathbf{P}_{\mathbf{z}\mathbf{z},i,k}^{-1} \mathbf{e}_{i,k} \sim \chi_{n_z}^2, \quad (22)$$

where $\chi_{n_z}^2$ is a chi-square distribution with a degree of freedom of n_z . The surprisal threshold β_k in (21) should be calculated in such a way that on average, $|l_k|$ cameras are selected as surprisal cameras. Thus,

$$\begin{aligned} \mathbb{E}\left[\sum_{i=1}^{|C_k|} s_{i,k}\right] &= \sum_{i=1}^{|C_k|} \mathbb{E}[s_{i,k}] \\ &= |C_k| \mathbb{E}[s_{i,k}] = |l_k|. \end{aligned} \quad (23)$$

From (21), (23), and (22), it is implied that

$$\Pr \left(\mathbf{e}_{i,k}^T \mathbf{P}_{\mathbf{z},i,k}^{-1} \mathbf{e}_{i,k} \geq \beta_k \right) = \frac{|l_k|}{|C_k|}. \tag{24}$$

The surprisal threshold β_k can be calculated as the value for which the probability of chi-square distributed squared and normalized measurement residual $\chi_{n_z}^2$ is greater than or equal to $|l_k| / |C_k|$ as

$$\beta_k = F_{\chi_{n_z}^2}^{-1} \left(1 - \frac{|l_k|}{|C_k|} \right), \tag{25}$$

where $F_{\chi_{n_z}^2}$ is the cumulative distribution function of the chi-square distribution $\chi_{n_z}^2$ with a degree of freedom of n_z .

Hence, the surprisal threshold β_k at time k can be calculated by using the knowledge of the number of cameras in the observation cluster $|C_k|$ and the number of desirable surprisal cameras $|l_k|$. Thus, the cameras c_i in the cluster can independently decide whether their local observations are informative or not.

5 Multi-camera object tracking with surprisal cameras (MOTSC)

In the proposed scheme, the cameras c_i where $i \in \{1, 2, \dots, M\}$ in the network that can observe an object at time k form a cluster (observation cluster) C_k with a FC $c_{o,k}$ as shown in the Fig. 2. The dynamic clustering can be achieved in several ways. One of such methods is pre-

sented in [39]. Further, each camera in the VSN has an on-board CIF algorithm. At each time k , each camera in the observation cluster C_k except the FC independently decides whether it is a surprisal camera or not, as discussed in Section 4. All surprisal cameras in the cluster C_k transmit their information contribution vectors and matrices to the FC. Moreover, the FC also performs the local filtering based on the on-board CIF. The locally calculated and received information contribution metrics are then fused together to achieve the estimated global state of the object at time k .

The FC is initialized with the global initial information vector and matrix $(\hat{\mathbf{y}}_{0|0}, \mathbf{Y}_{0|0})$. At each time step k , it has four main functions: surprisal threshold calculation, local filtering, information fusion, and global state dissemination as shown in Algorithm 1.

- *Surprisal threshold calculation:* The surprisal threshold can be calculated with the knowledge of the size $|C_k|$ of the observation cluster and desired size $|l_k|$ of the surprisal cluster as shown in (25). Hence, the FC which knows this information calculates and broadcasts the surprisal threshold whenever the observation and surprisal cluster sizes change.
- *Local filtering:* The FC performs the local estimation based on its measurement $\mathbf{z}_{o,k}$ by using the on-board CIF. Firstly, the FC predicts the information vector and matrix $(\hat{\mathbf{y}}_{o,k|k-1}, \mathbf{Y}_{o,k|k-1})$ from the prior global information vector and matrix $(\hat{\mathbf{y}}_{k-1|k-1}, \mathbf{Y}_{k-1|k-1})$

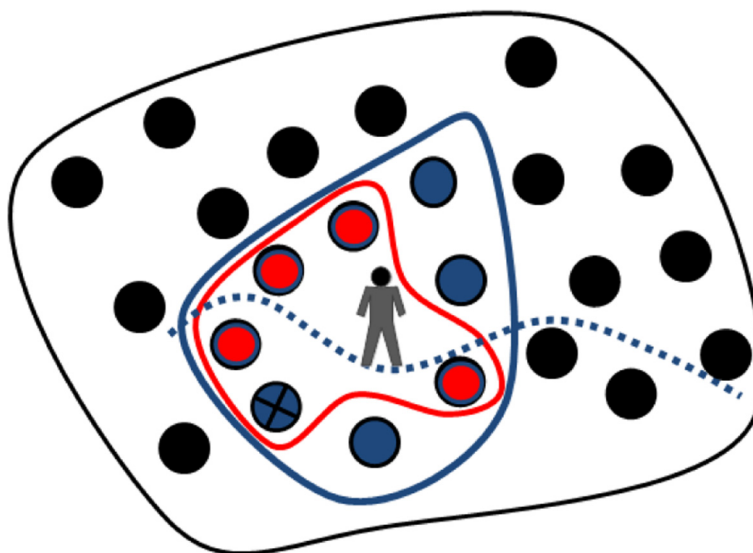


Fig. 2 The VSN with the observation and surprisal clusters with a fixed FC. All the cameras (dots) inside the blue cluster can observe the object at a given time and form the observation cluster. The cameras (dots) inside the red cluster are the surprisal cameras and form the surprisal cluster. The blue dot with cross represents the FC

as shown in Appendix 1. Then, it computes the information contribution vector and matrix $(\mathbf{i}_{o,k}, \mathbf{I}_{o,k})$ by using its own local observations $\mathbf{z}_{o,k}$ as shown in Appendix 2.

- **Information fusion:** The FC receives a set of information contribution metrics $(\mathbf{i}_{i,k}, \mathbf{I}_{i,k})$ where $i = 1, 2, \dots, |l_k|$ from the surprisal cameras in the cluster. The global information vector and information matrix $(\hat{\mathbf{y}}_{k|k}, \mathbf{Y}_{k|k})$ are obtained by fusing the received surprisal information contributions and its own information contributions $(\mathbf{i}_{f,k}, \mathbf{I}_{f,k})$ with the predicted information vector and matrix $(\hat{\mathbf{y}}_{k|k-1}, \mathbf{Y}_{k|k-1})$.
- **Global state dissemination:** After the information vector and matrix $(\hat{\mathbf{y}}_{k|k}, \mathbf{Y}_{k|k})$ are computed, the FC broadcasts it in the network. Hence, the cameras in the network have the global knowledge which can be used as prior information for the local filtering in the time step $k + 1$.

The cameras in the observation cluster C_k at time k have two main functions to perform: time update and surprisal update as shown in Algorithm 2. The cameras in the observation cluster know the prior global information of the object $(\hat{\mathbf{y}}_{k-1|k-1}, \mathbf{Y}_{k-1|k-1})$. At each time step k , they perform the following:

- **Time update:** The camera predicts the information vector and matrix $(\hat{\mathbf{y}}_{i,k|k-1}, \mathbf{Y}_{i,k|k-1})$ from the prior global information vector and matrix $(\hat{\mathbf{y}}_{k-1|k-1}, \mathbf{Y}_{k-1|k-1})$ using the CIF time update as shown in Appendix 1.
- **Surprisal update:** Each camera receives the surprisal threshold β_k from the FC whenever the observation and/or surprisal cluster size changes. Upon receiving the measurement $\mathbf{z}_{i,k}$, each camera c_i calculates the corresponding measurement residual and innovation covariance $(\mathbf{e}_{i,k}, \mathbf{P}_{zz,i,k})$. The proposed surprisal threshold rule in Section 4.3 is used to determine whether it is a surprisal camera or not. If the camera is a surprisal camera, the information contribution vector and matrix $(\mathbf{i}_{i,k}, \mathbf{I}_{i,k})$ are calculated according to (9) and (10). Thereafter, the information metrics are transmitted to the FC. If the camera is not a surprisal camera, then the surprisal update is aborted.

After the surprisal update, each camera c_i in the network receives the global information $(\hat{\mathbf{y}}_{k|k}, \mathbf{Y}_{k|k})$ from the FC. Hence, each camera in the network has the knowledge of the global state of the object which can also be used as the prior information in the local estimation for the next time step $k + 1$.

In this paper, the FC is assumed to be fixed and not effected by node failures. It is also assumed that

the delays in transmitting local information to the FC are all less than the sampling interval of the cameras. Thus, the FC can fuse the arriving information contribution in time. The communication links in the network are assumed to be perfect. Hence, the only cause of a missing information metric from a camera is that the corresponding observations are not informative enough.

Algorithm 1: MOTSC at Fusion Center (FC)

Start: Initialize the filter with $(\hat{\mathbf{y}}_{0|0}, \mathbf{Y}_{0|0})$.

At each time k , the prior information vector and matrix $(\hat{\mathbf{y}}_{k-1|k-1}, \mathbf{Y}_{k-1|k-1})$ are known

1. Surprisal Threshold Calculation

if $(|l_k|, |C_k|) \neq (|l_{k-1}|, |C_{k-1}|)$ **then**

(a) Calculate the new surprisal threshold β_k
/* Equation (25) */

(b) Broadcast β_k in the observation cluster

2. Local Filtering

(a) Compute the predicted information vector and matrix based on the CIF time update

$[\hat{\mathbf{y}}_{k|k-1}, \mathbf{Y}_{k|k-1}] = \text{Time}$

Update $[\hat{\mathbf{y}}_{k-1|k-1}, \mathbf{Y}_{k-1|k-1}]$ /* Appendix 1 */

(b) Upon receiving the measurement $\mathbf{z}_{o,k}$, calculate the information contribution vector and matrix based on the CIF measurement update

$[\mathbf{i}_{f,k}, \mathbf{I}_{f,k}] = \text{Measurement Update}$

$[\hat{\mathbf{y}}_{k|k-1}, \mathbf{Y}_{k|k-1}, \mathbf{z}_{f,k}]$ /* Appendix 2 */

3. Information Fusion

(a) Receive the information contribution metrics $\{\mathbf{i}_{i,k}, \mathbf{I}_{i,k}\}$, where $i \in C_k$ from the surprisal cameras in the cluster

(b) Perform information fusion to achieve the global information

$[\hat{\mathbf{y}}_{k|k}, \mathbf{Y}_{k|k}] = \text{Information Fusion}$

$[\hat{\mathbf{y}}_{k|k-1}, \mathbf{Y}_{k|k-1}, \{\mathbf{i}_{i,k}, \mathbf{I}_{i,k}\}, (\mathbf{i}_{f,k}, \mathbf{I}_{f,k})]$

4. Global State Dissemination

Broadcast global information $[\hat{\mathbf{y}}_{k|k}, \mathbf{Y}_{k|k}]$ in the network

Algorithm 2: MOTSC at each local camera in the cluster

Start: Initialize the filter at each camera with $(\hat{\mathbf{y}}_{0|0}, \mathbf{Y}_{0|0})$
 At each time k , each camera receives the prior information vector and matrix $(\hat{\mathbf{y}}_{k-1|k-1}, \mathbf{Y}_{k-1|k-1})$ from the FC.

1. Time Update

Compute the predicted information vector and matrix based on the CIF time update
 $[\hat{\mathbf{y}}_{i,k|k-1}, \mathbf{Y}_{i,k|k-1}] = \text{Time}$
 Update $[\hat{\mathbf{y}}_{k-1|k-1}, \mathbf{Y}_{k-1|k-1}]$ /* Appendix 1 */

2. Surprisal Update

(a) Upon receiving the measurement $\mathbf{z}_{i,k}$, calculate the measurement residual $\mathbf{e}_{i,k}$
 /* steps 1 to 4 in Appendix 2 */

(b) Calculate the innovation covariance $\mathbf{P}_{\mathbf{z}\mathbf{z},i,k}$
 (c) Compare the information content in the measurement residual $\mathbf{e}_{i,k}$ with surprisal threshold β_k

if $(\mathbf{e}_{i,k}^T \mathbf{P}_{\mathbf{z}\mathbf{z},i,k}^{-1} \mathbf{e}_{i,k} \geq \beta_k)$ **then**

I Calculate the information contribution vector and matrix $(\mathbf{i}_{i,k}, \mathbf{I}_{i,k})$
 /* steps 5 to 7 in Appendix 2 */

II Transmit $(\mathbf{i}_{i,k}, \mathbf{I}_{i,k})$ to the FC

6 Results

In this section, the efficiency of the proposed MOTSC method is evaluated based on the simulation and experimental data. In our approach, the efficiency is defined in terms of the sum of the root mean square errors (RMSEs) of the estimated global state and the ground truth in x and y directions. Moreover, the energy and bandwidth efficiency are calculated in terms of the average number of transmissions from the cameras in the observation cluster to the FC.

6.1 Simulation results

The simulation considers a VSN with cameras having overlapping FOVs as shown in Fig. 2. All of the cameras that can observe the xy -plane, where $x \in [-500, 500]$ and $y \in [-500, 500]$ form an observation cluster with a FC. The motion of the object is modeled with Gaussian distributed acceleration as given in (1). The ground truth of the position of the object is simulated by assuming that the process noise covariance \mathbf{Q}_k and measurement noise

covariance \mathbf{R}_k are $\text{diag}(5, 5)$ and $\text{diag}(1, 1)$, respectively. Each camera c_i in the cluster has its own homography function h_i . Since we assume static cameras, the homography of the cameras do not change with time k and object. The algorithms are evaluated on 1000 different trajectories with different initializations. Figure 3 shows some of the simulated trajectories of the object.

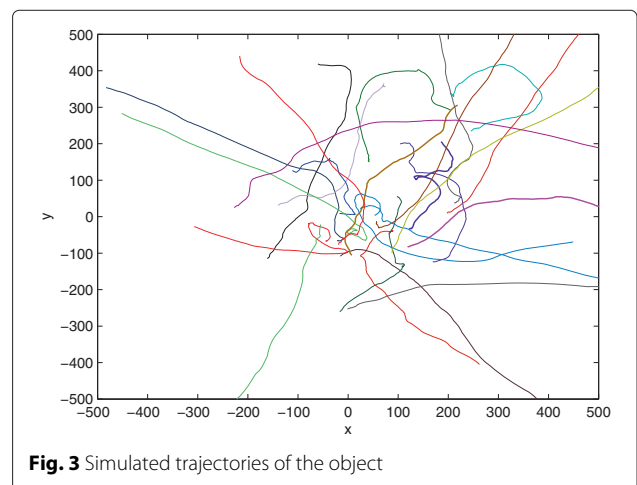
6.1.1 Scenario 1

In this scenario, the accuracy of the CIF- and EIF-based object tracking methods in the VSN are compared. In this comparison, the proposed surprisal selection method is not employed. Hence, all the cameras in the observation cluster participate in the information fusion. In the abovementioned simulation setup, each camera calculates the local information metrics based on the local observations. The information metrics from the local cameras are fused at the FC. Moreover, the process noise covariance \mathbf{Q}_k and measurement noise covariance \mathbf{R}_k are considered to be known to all the cameras in the cluster. The cluster is also assumed to be fully connected with perfect communication links to the FC.

Under the above conditions, Fig. 4 shows the average RMSE (ARMSE) of the multi-camera object tracking methods based on the CIF and EIF for different observation cluster sizes. To achieve statistical reliability, the RMSE is averaged over a thousand simulation runs and 1000 simulated trajectories to yield the ARMSE. From Fig. 4, we can infer that the CIF-based object tracking outperforms the EIF-based method, though the tracking accuracy of the two methods improves with increasing cluster size.

6.1.2 Scenario 2

In this scenario, the accuracy of the proposed MOTSC is analyzed in comparison with multi-camera object



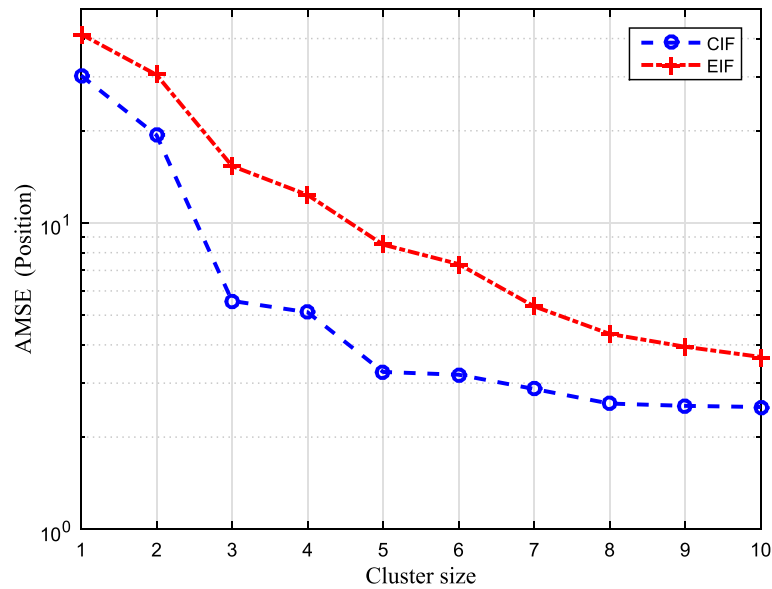


Fig. 4 Tracking accuracy of the multi-camera object tracking methods based on the CIF and EIF in terms of average RMSE for different cluster sizes

tracking with random cameras, fixed cameras, best cameras, and active sensing cameras.

- *Multi-camera object tracking with random cameras (MOTRC)*: A random subset of cameras in the observation cluster transmit their local information metrics to the FC independent of the information contained in their measurements.
- *Multi-camera object tracking with fixed cameras (MOTFC)*: A fixed subset of cameras in the observation cluster transmit their local information metrics to the FC.
- *Multi-camera object tracking with best cameras (MOTBC)*: All the cameras in the observation cluster C_k send their surprisal of the measurement residual to the FC. The FC ranks the cameras in the ascending order of their surprisal score and informs $|I_k|$ best cameras to share their local information metrics. Then, the informed cameras send their local information metrics to the FC. The total number of transmissions to and from the FC involved in this method are $\|C_k\| + 2|I_k|$. The MOTBC method is an adoption from [36].
- *Multi-camera object tracking method with active sensing cameras (MOTAC)*: The FC activates or deactivates the cameras from participating in information exchange by maximizing reward-cost utility function as given in [35]. The reward is expected information gain (EIG). At each time k , the FC evaluates the utility function for all possible activated and deactivated camera combinations

before activating the best cameras to participate in the information fusion. Refer to [35] for complete details.

Figure 5 shows the RMSE of the MOTSC, MOTRC, MOTFC, MOTBC, and MOTAC methods. The x -axis of the figure represents the average number of cameras participated in the information fusion at each time k . The total number of cameras $|C_k|$ in the observation cluster remains 10. From Fig. 5, we can infer that the tracking accuracy of these methods improves with increasing size of the subset that can participate in the information fusion. However, the proposed MOTSC method outperforms both the MOTRC and MOTFC for the same number of cameras $|I_k|$ that can transmit to the FC. The MOTSC, MOTBC, and MOTAC methods approximately achieve the same tracking accuracy. However, in the MOTAC method, at each time k , the FC has to evaluate the reward-cost utility function for all possible activated and deactivated camera combinations (2^{10} in this case) before selecting the best possible cameras to participate in the information fusion. Moreover, the camera selection at time k in the MOTAC method does not depend on the current measurements. In the MOTBC method, in order to select the best possible cameras, the FC has to receive the surprisal scores from all the cameras in the observation cluster. The centralized and complex camera selection restricts the scalability of both the MOTAC and MOTBC methods. On the other hand, in the proposed MOTSC method, the cameras take decision independently whether to participate in information fusion or not.

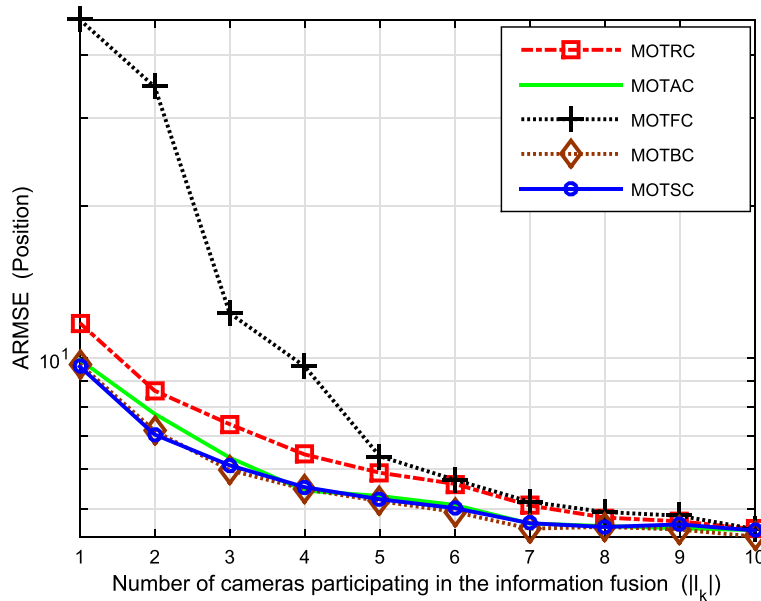


Fig. 5 Tracking accuracy of the MOTSC, MOTRC, MOTFC, MOTBC, and MOTAC methods. The size $|C_k|$ of the observation cluster is 10, and the size of the fixed, random, or surprisal subset varies from 1 to 10

On the other hand, Fig. 6 shows the number of transmissions sent to the FC in the MOTSC and MOTRS methods. The x -axis shows the theoretical number $|I_k|$ of surprisal cameras which is used to calculate the surprisal threshold. The y -axis shows the number of transmissions to the FC from the surprisal and random cameras in the corresponding methods. From the figure, it is illustrated that on

average, the number of transmissions to the FC for both methods is approximately equal and matches the theoretical requirements. Even though the MOTBC achieves the same performance as the MOTSC, the number of transmissions in MOTBC is equal to $|C_k| + 2|I_k|$ which can be significantly higher than the average number of transmissions $|I_k|$ in MOTSC.

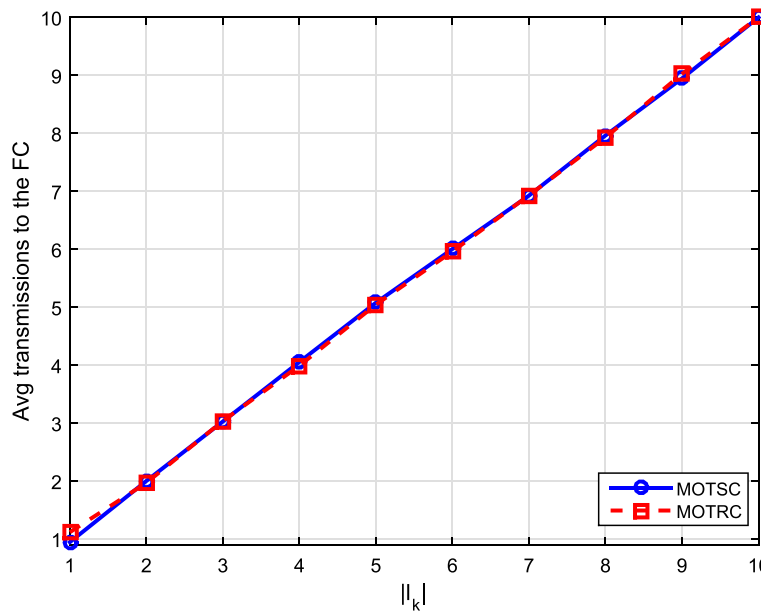


Fig. 6 The average number of transmissions to the FC in the MOTSC and MOTRC methods. The observation cluster size C_k is 10, and the size $|I_k|$ of the random or surprisal subset varies from 1 to 10

6.2 Experimental results

The experimental setup consists of a self-aware multi-camera cluster built in the lab of our institute. The camera cluster consists of four atom-based cameras (1.6 GHz processor, 2 GB RAM, 30 GB internal SSD hard disk) from SLR Engineering and two PandaBoards on which the middle-ware system ELLA [40] is developed. The cameras in the cluster can perform object detection and tracking together with state estimation locally. Moreover, they are connected via Ethernet. In the experimental setup, the four cameras in the network have overlapping FOVs. The motion of the object is modeled by predefined tracks. The experiment considers ten different such predefined tracks within the overlapping FOV of the four cameras. Figure 7 shows some of the object tracks that are used for evaluating the proposed MOTSC method. The x - and y -axes represent the dimensions of the lab where the experimental setup is built. Each track has a duration of 120 s. Each camera c_i in the cluster has its own homography function h_i . Since we assume fixed cameras, the homography of the cameras does not change with time k . The process noise covariance \mathbf{Q}_k and measurement noise covariance \mathbf{R}_k are considered as $\text{diag}(10, 10)$ and $\text{diag}(2, 2)$, respectively.

Figure 8 shows the average RMSE of the MOTSC, MOTBC, and MOTRC methods. The x -axis of the figure represents the size $|I_k|$ of random and surprisal subset of the cameras that transmit their local information metrics to the FC at each time k . The total number of cameras $|C_k|$ in the observation cluster remains four irrespective of the desired size of the random and surprisal subset. To achieve statistical reliability, the average RMSE is averaged over ten predefined tracks discussed above. From Fig. 8, we can infer that the proposed MOTSC outperforms the MOTRC for the same number of cameras $|I_k|$ that can participate in the information fusion. Even though the MOTBC method achieves approximately the same

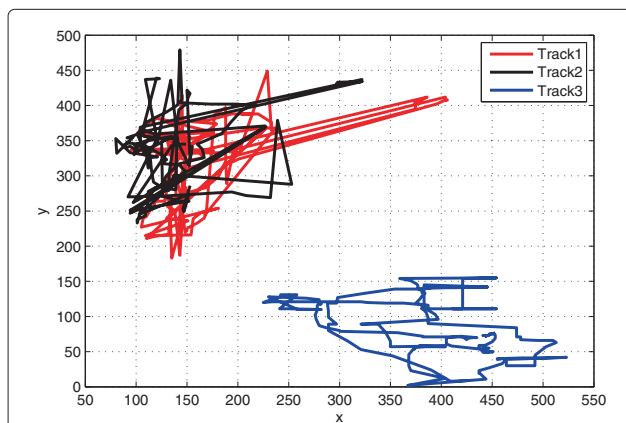


Fig. 7 Examples of the predefined object tracks used for evaluating the MOTSC method

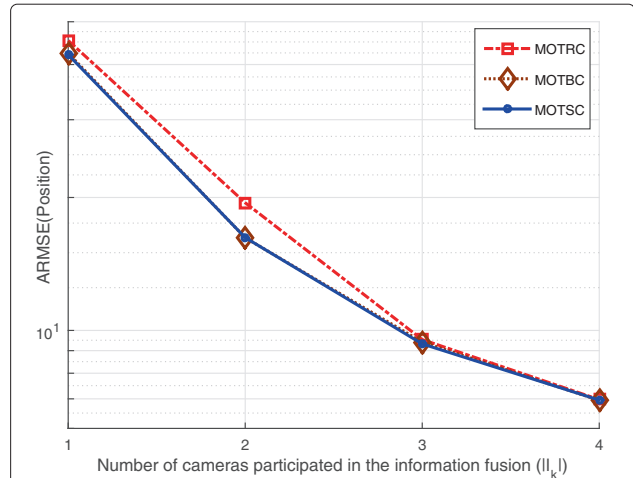


Fig. 8 Tracking accuracy of the MOTSC, MOTRC, and MOTBC methods in the experimental setup defined above. The size of the observation cluster is 4, and the size $|I_k|$ of the random or surprisal subset varies from 1 to 4

tracking accuracy as the MOTSC method, the number of transmissions to the FC is always $|C_k| + 2|I_k|$.

On the other hand, Fig. 9 shows the average number of transmissions sent to the FC in the MOTSC and MOTRC methods. The x -axis shows the theoretical number $|I_k|$ of surprisal cameras which is used to calculate the surprisal threshold. The y -axis shows the average number of transmissions to the FC by the corresponding methods during the experiment. From the figure, it is illustrated that on average, the number of transmissions to the FC for both the methods is approximately equal and matches the theoretical requirements. Hence, the proposed MOTSC shows better accuracy than that of the MOTRC for the same number of average transmissions.

7 Conclusions

In this work, a multi-camera object tracking with surprisal cameras in a VSN is proposed. The cameras in the VSN that can observe an object form an observation cluster with a fixed FC. However, due to bandwidth constraints and energy limitations, it is usually desirable to have only a subset of cameras exchanging their local information to the fusion center. In our approach, each camera runs a local object tracking algorithm based on the on-board CIF. Each camera independently determines whether its observations are informative enough or not by using the surprisal of its measurement residual. Only if a camera's measurements are informative enough (surprisal cameras), it calculates and transmits the local information vector and matrix to the fusion center. The global state of the object is obtained by fusing the local information from surprisal cameras at the fusion center. The proposed scheme also ensures that on average, only a

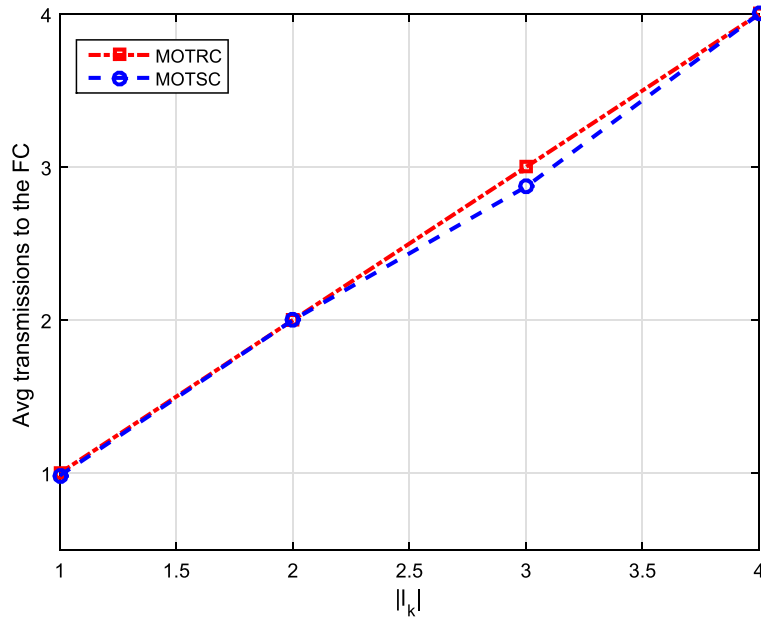


Fig. 9 The average of number of transmissions sent to the FC in the multi-camera object tracking with surprisal and random cameras. The observation cluster size $|C_k|$ is 4, and the size $|I_k|$ of the random or surprisal subset varies from 1 to 4

desired number of cameras participate in the information exchange. The proposed multi-camera object tracking with surprisal cameras shows a considerable improvement in tracking accuracy over the multi-camera object tracking with random and fixed cameras for the same number of transmissions to the fusion center.

Endnote

¹In general, the surprisal is defined for the discrete random variables (DRV). Hence, we are considering the innovation to be a DRV.

Appendices

The multi-sensor CIF constitutes of three main steps: time update and measurement update at each sensor i and time k .

Appendix 1: time update (TU)

Calculate the predicted information vector and information matrix $[\hat{\mathbf{y}}_{i,k|k-1}, \mathbf{Y}_{i,k|k-1}]$ from global prior information $[\hat{\mathbf{y}}_{k-1|k-1}, \mathbf{Y}_{k-1|k-1}]$.

1. Calculate the state estimate

$$\hat{\mathbf{x}}_{k-1|k-1} = \mathbf{Y}_{k-1|k-1} \hat{\mathbf{y}}_{k-1|k-1}$$

2. Compute the cubature points $m = (1, 2, \dots, 2n_x)$

$$\mathbf{cp}_{m,k-1|k-1} = \sqrt{\mathbf{Y}_{k-1|k-1}^{-1}} \xi_m + \hat{\mathbf{x}}_{k-1|k-1}$$

where n_x is the length of the state vector. ξ_m represent the m th intersection point of the surface of the n -dimensional unit sphere and its axes.

3. Propagate the cubature points through the motion model

$$\mathbf{x}_{m,k|k-1}^* = \mathbf{f}_{i,k}(\mathbf{cp}_{m,i,k-1|k-1})$$

4. Calculate the predicted state as

$$\hat{\mathbf{x}}_{i,k|k-1} = \frac{1}{2n_x} \sum_{m=1}^{2n_x} \mathbf{x}_{m,i,k|k-1}^*$$

5. Calculate the predicted error covariance as

$$\mathbf{P}_{k|k-1} = \mathbf{M}_{i,k|k-1} \mathbf{M}_{i,k|k-1}^T + \mathbf{Q}_{i,k}^s$$

where $\mathbf{Q}_{i,k}$ is the process noise covariance. The predicted weighted centered matrix $\mathbf{M}_{i,k|k-1}$ is given as

$$\mathbf{M}_{i,k|k-1} = \frac{1}{\sqrt{2n}} \begin{bmatrix} \mathbf{x}_{1,i,k|k-1}^* - \hat{\mathbf{x}}_{i,k|k-1} & \mathbf{x}_{2,i,k|k-1}^* - \hat{\mathbf{x}}_{i,k|k-1} \\ \vdots & \vdots \\ \mathbf{x}_{2n,i,k|k-1}^* - \hat{\mathbf{x}}_{i,k|k-1} \end{bmatrix}$$

6. Compute the predicted information matrix and predicted information vector

$$\mathbf{Y}_{i,k|k-1} = \mathbf{P}_{i,k|k-1}^{-1}$$

$$\hat{\mathbf{y}}_{i,k|k-1} = \mathbf{Y}_{i,k|k-1} \hat{\mathbf{x}}_{i,k|k-1}$$

Appendix 2: measurement update (MU)

Each sensor calculates its information contribution vector and matrix $[\mathbf{i}_{i,k}, \mathbf{I}_{i,k}]$ from the predicted information

vector and matrix $[\hat{\mathbf{y}}_{i,k|k-1}, \mathbf{Y}_{i,k|k-1}]$ and the measurement $\mathbf{z}_{i,k}$.

1. Calculate the cubature points

$$\mathbf{c}\mathbf{p}_{m,i,k|k-1} = \sqrt{\mathbf{P}_{i,k|k-1}\xi_m} + \hat{\mathbf{x}}_{i,k|k-1}.$$

2. Propagate the cubature points through the observation function

$$\mathbf{z}_{m,i,k|k-1}^* = \mathbf{h}_{i,k}(\mathbf{c}\mathbf{p}_{m,i,k|k-1}).$$

3. Calculate the predicted measurement

$$\hat{\mathbf{z}}_{i,k|k-1} = \frac{1}{2n_x} \sum_{m=1}^{2n_x} \mathbf{z}_{m,i,k|k-1}^*.$$

4. Calculate the measurement residual

$$\mathbf{e}_{i,k} = \mathbf{z}_{i,k} - \hat{\mathbf{z}}_{i,k|k-1}.$$

5. Calculate the cross covariance

$$\mathbf{P}_{\mathbf{z}\mathbf{x},i,k|k-1} = \frac{1}{2n} \sum_{m=1}^{2n} \mathbf{c}\mathbf{p}_{m,i,k|k-1} \mathbf{z}_{m,i,k|k-1}^{*T} - \hat{\mathbf{x}}_{i,k|k-1} \hat{\mathbf{z}}_{i,k|k-1}^T.$$

6. Calculate the information contribution matrix

$$\mathbf{I}_{i,k} = \mathbf{Y}_{i,k|k-1} \mathbf{P}_{\mathbf{z}\mathbf{x},i,k|k-1} \mathbf{R}_{i,k}^{-1} \mathbf{P}_{\mathbf{z}\mathbf{x},i,k|k-1}^T \mathbf{Y}_{i,k|k-1}^T,$$

where $\mathbf{R}_{i,k}$ is the measurement noise covariance matrix.

7. Compute the information contribution vector

$$\mathbf{i}_{i,k} = \mathbf{Y}_{i,k|k-1} \mathbf{P}_{\mathbf{z}\mathbf{x},i,k|k-1} \mathbf{R}_{i,k}^{-1} (\mathbf{e}_{i,k} + \mathbf{P}_{\mathbf{z}\mathbf{x},i,k|k-1}^T \mathbf{Y}_{i,k|k-1}^T \hat{\mathbf{x}}_{i,k|k-1}).$$

Competing interests

The authors declare that they have no competing interests.

Acknowledgements

This work was supported in part by the EACEA Agency of the European Commission under EMJD ICE FPA no. 2010-0012. The work has also been supported in part by the ERDF, KWF, and BABEG under grant KWF-20214/21530/32602 (ICE Booster). It has been performed in the research cluster Lakeside Labs.

Author details

¹Institute of Networked and Embedded Systems, Alpen-Adria-Universität, Klagenfurt, Austria. ²Department of Marine Engineering, Electrical, Electronics, and Telecommunications, University of Genova, Genova, Italy. ³Institute of Signal Processing, Johannes Kepler University, Linz, Austria.

Received: 9 August 2015 Accepted: 7 April 2016

Published online: 26 April 2016

References

1. B Rinner, M Quaritsch, W Schriebl, T Winkler, W Wolf, *The evolution from single to pervasive smart cameras. Paper presented at the 2nd ACM/IEEE international conference on distributed smart cameras.* (IEEE, Stanford, CA, 2008), pp. 1–10
2. S Soro, W Heinzelman, A survey of visual sensor networks. *Adv. Multimed* (2009)
3. A Yilmaz, O Javed, M Shah, Object tracking: a survey. *ACM Comput. Surv.* **38**(4), 1–45 (2006)
4. S-K Weng, C-M Kuo, S-K Tu, Video object tracking using adaptive Kalman filter. *J. Visual Commun. Image Represent.* **17**(6), 1190–1208 (2006)
5. H Wang, D Suter, K Schindler, C Shen, Adaptive object tracking based on an effective appearance filter. *IEEE Trans. Pattern Anal. Mach. Intell.* **29**(9), 1661–1667 (2007)
6. W-J Liu, Y-J Zhang, Edge-colour-histogram and Kalman filter-based real-time object tracking. *J. Tsinghua Univ. (Sci. Technol.)* **48**(7) (2008)
7. R Olfati-Saber, NF Sandell, Distributed tracking in sensor networks with limited sensing range. *Am. Control Conf.* 3157–3162 (2008)
8. C Soto, S Bi, AK Roy-Chowdhury, Distributed multi-target tracking in a self-configuring camera network. *IEEE Conf. Comput. Vis. Pattern Recogn.* 1486–1493 (2009)
9. H-M Wang, L-L Huo, J Zhang, Target tracking algorithm based on dynamic template and Kalman filter. *IEEE Int. Conf. Commun. Softw. Netw.* 330–333 (2011)
10. B Song, C Ding, AT Kamal, JA Farrell, AK Roy-Chowdhury, Distributed camera networks. *IEEE Signal Process. Mag.* **28**(3), 20–31 (2011)
11. SY Chen, Kalman filter for robot vision: a survey. *IEEE Trans. Ind. Electron.* **59**(11), 4409–4420 (2012)
12. R Rosales, S Sclaroff, Improved tracking of multiple humans with trajectory prediction and occlusion modeling. *IEEE CVPR Workshop Int. Vis. Motion* (1998)
13. P Li, T Zhang, B Ma, Unscented Kalman filter for visual curve tracking. *Image and Vision Comput.* **22**(2), 157–164 (2004)
14. M Meuter, U Iurleg, S-B Park, A Kummert, The unscented Kalman filter for pedestrian tracking from a moving host. *IEEE Intell. Veh. Symp.* 37–42 (2008)
15. VP Bhuvana, M Schranz, M Huemer, B Rinner, Distributed object tracking based on cubature Kalman filter. *Asilomar Conf. Signals, Syst. Comput.* 423–427 (2013)
16. K Nummiaro, E Koller-Meier, L Van Gool, An adaptive color-based particle filter, image and vision computing. **21**(1), 99–110 (2003)
17. K Okuma, A Taleghani, N Freitas, JJ Little, DG Lowe, A boosted particle filter: multitarget detection and tracking. *Eur. Conf. Comput. Vis* (2004)
18. Y Rui, Y Chen, Better proposal distributions: object tracking using unscented particle filter. *IEEE Comput. Soc. Conf. Comput. Vis Pattern Recognit.* **2**, 786–793 (2001)
19. C-C Wang, C Thorpe, S Thrun, M Hebert, H Durrant-Whyte, Simultaneous localization, mapping and moving object tracking. *Int. J. Robot. Res.* **26**, 889–916 (2007)
20. Y Rathi, N Vaswani, A Tannenbaum, A Yezzi, Tracking deforming objects using particle filtering for geometric active contours. *IEEE Transactions on Pattern Analysis and Machine Intelligence.* **29**(8), 1470–1475 (2007)
21. Y Li, H Ai, T Yamashita, S Lao, M Kawade, Tracking in low frame rate video: a cascade particle filter with discriminative observers of different life spans. *IEEE Trans. Pattern Anal. Mach. Intell.* **30**(10), 1728–1740 (2008)
22. MD Breitenstein, F Reichlin, B Leibe, E Koller-Meier, LV Gool, Robust tracking-by-detection using a detector confidence particle filter. *IEEE Int. Conf. Comput. Vis.* 1515–1522 (2009)
23. AD Bimbo, F Dini, Particle filter-based visual tracking with a first order dynamic model and uncertainty adaptation. *Comput. Vis. Image Underst.* **115**(6), 771–786 (2011)
24. Z Ni, S Sunderrajan, A Rahimi, BS Manjunath, Distributed particle filter tracking with online multiple instance learning in a camera sensor network. *17th IEEE Int. Conf. Image Process.* 37–40 (2010)
25. C Ding, B Song, AA Morye, JA Farrell, AKR Chowdhury, Collaborative sensing in a distributed PTZ camera network. *IEEE Trans. Image Process.* **21**(7), 3282–3295 (2012)
26. AT Kamal, JA Farrell, AK Roy-Chowdhury, Consensus-based distributed estimation in camera networks. *IEEE Int. Conf. Image Process.* 1109–1112 (2012)
27. H Medeiros, J Park, AC Kak, Distributed object tracking using a cluster-based Kalman filter in wireless camera networks. *IEEE J. Sel. Top. Signal Process.* **2**(4), 448–463 (2008)
28. S Dan, *Optimal State Estimation: Kalman, H Infinity, and Nonlinear Approaches.* (John Wiley & Sons, 2006)
29. AT Kamal, C Ding, B Song, JA Farrell, AK Roy-Chowdhury, A generalized Kalman consensus filter for wide-area video networks. *IEEE Conf. Decis. Control. Eur. Control.* 7863–7869 (2011)

30. AT Kamal, JA Farrell, AK Roy-Chowdhury, Information weighted consensus. *IEEE Annu. Conf. Decis. Control* (2012)
31. VP Bhuvana, M Huemer, CS Regazzoni, Distributed object tracking based on square root cubature H-infinity information filter. *IEEE Int. Conf. Inf. Fusion*, 1–6 (2014)
32. K Morioka, K Szilveszter, J-H Lee, P Korondi, H Hashimoto, A cooperative object tracking system with fuzzy-based adaptive camera selection. *Int. J. smart Sens. Intell. Syst.* **3**, 338–58 (2010)
33. DB Yang, J Shin, AO Ercan, LJ Guibas, Sensor tasking for occupancy reasoning in a network of cameras. *Stanf. Netw. Res. Center* (2010)
34. L Tessens, M Morbee, H Aghajan, W Philips, Camera selection for tracking in distributed smart camera networks. *ACM Trans. Sensor Netw.* **10**, 1–33 (2014)
35. A de San Bernabe, JR Martinez-de Dios, A Ollero, Entropy-aware cluster-based object tracking for camera wireless sensor networks. *IEEE/RSJ Int. Conf. Intell. Robot. Syst.* 3985–3992 (2012)
36. E Shen, R Hornsey, in *Proceedings of the 5th ACM/IEEE International Conference on Distributed Smart Cameras*. Local image quality metric for a distributed smart camera network with overlapping FOVs, (2011), pp. 1–6
37. CE Shannon, A mathematical theory of communications. *Bell Syst. Technical J.* **27**, 379–423 (1948)
38. I Arasaratnam, S Haykin, Cubature Kalman filters. *IEEE Trans. Auto. Control.* **54**(6), 1254–1269 (2009)
39. M Schranz, B Rinner, Resource-aware state estimation in visual sensor networks with dynamic clustering. *4th Int. Conf. Sensor Netw.* 10 (2015)
40. B Dieber, J Simonjan, L Esterle, B Rinner, G Nebehay, R Pflugfelder, GJ Fernandez, Ella: Middleware for multi-camera surveillance in heterogeneous visual sensor networks. *ACM/IEEE Int. Conf. Distrib. Smart Cameras*, 1–6 (2013)

Submit your manuscript to a SpringerOpen[®] journal and benefit from:

- Convenient online submission
- Rigorous peer review
- Immediate publication on acceptance
- Open access: articles freely available online
- High visibility within the field
- Retaining the copyright to your article

Submit your next manuscript at ► springeropen.com
