Camera Calibration from Multi-Object Triangulation and Tracking using Second-Order Moment Statistics

Èric Pairet, Jose Bernal, Isabel Schlangen and Daniel E. Clark
School of Electrical and Physical Sciences, Heriot-Watt University, Edinburgh, EH14 4AS, UK
Email: {ep18, jab2, is117, D.E.Clark}@hw.ac.uk

***Abstract*—Triangulating objects from two cameras has been widely addressed in computer vision. Classical approaches consider establishing the relation between the view of the two cameras from static setups, e.g. patterns placed in the Field of View (FoV) of the sensing system. Nowadays, the same task can be achieved dynamically by triangulating and tracking moving objects while calibrating the cameras at the same time. For instance, the multi-object estimation could be conditioned on the states of the cameras which are given by a parent process optimising them. This process can be feedback with the variance in the number of objects in the scene as a measure of confidence in the estimation of the camera parameters. In this paper, we present an approach for integrating the calculation of variance in a unified framework for camera calibration from multiple object triangulation and tracking. The proposed integration is evaluated in simulated and real environments.**

***Index Terms*—Camera calibration, disparity space, multi-object tracking, PHD filter, higher-order statistics.**

I. Introduction

Estimating the 3D position of an object of interest using two cameras is a well-known problem in the community of computer vision [1], [2], [3], [4], [5]. Note that for addressing this problem, knowledge of the geometry of the scene, as well as specific characteristics of the set of cameras (i.e. intrinsic and extrinsic parameters), is required in advance.

Given a pair of cameras, the initial step consists in determining the projective geometry between the two different views. This task is classically performed by finding the relation between the real world and two image planes in two steps. First, a pattern with some key-points is placed in the FoV of the two cameras. The real position of these key points in the real world is estimated beforehand. Second, the epipolar geometry is computed through calibration methods, such as the seven points [6], eight points [7], rank-2 constraint [8], among others. Note that the selection of the calibration method depends largely on the requirements of the specific case of study. For instance, the literature recognises that linear methods are useful in ideal environments in which no noise and no outliers are presented, while the so-called robust methods outperform in presence of both issues [9]. Third, the images are rectified [10], so that a similar setup to the one depicted in Fig. 1 is obtained. In this sense, finding the correspondence between a point in the image plane to another in the other one is reduced to a search on the corresponding epipolar line. Hereupon, the 3D pose estimation of objects is achieved by taking into account the information of the two cameras.

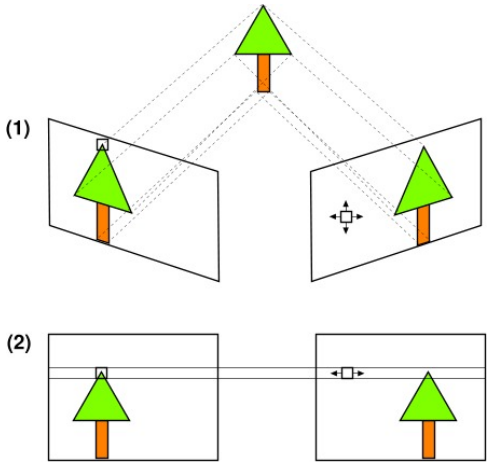

Fig. 1. Rectified and non-rectified views from two cameras [11]. In (1) the tree is projected onto two non-rectified cameras, while (2) depicts a rectified setup in which a point in the left image plane corresponds to an epipolar line in the other.

Nowadays, triangulation of an object from two cameras can be carried out dynamically, for example, by detecting and tracking paper planes thrown to the FoV of the sensing system as presented in [12]. In that work, the authors established a Bayesian framework for estimating and tracking the 3D position of the targets while calibrating the pair of cameras at the same time. Since the computation of the Bayesian filter is expensive, an approximation of this scheme through the propagation of the first-order moment of a multi-object distribution, also referred as Probability Hypothesis Density (PHD) [13], [14], is considered. As demonstrated by experiments on simulated and real data, the proposal was able to provide coherent results for the three tasks it unifies.

It has been shown recently that regional statistics, such as expectation and variance in the number of targets, can be calculated within multi-object filtering process [15]. Note that having a sense of variance in the number of targets might be useful for taking strategic decisions. For instance, this measure could be used for determining how reliable the information collected by the different sensors in the scene is.

The aim of this work is to introduce the concept of variance in the number of targets into the multi-object triangulation, tracking and camera calibration framework proposed by Houssineau *et al* [12]. The paper is organised as follows. In Section II, brief details of the considered framework are described. Then, in Section III, we discuss how the con-

cept of variance in the number of targets is plugged into the framework. To understand better where the variance is computed, the pseudo-code of the data update on the PHD filter is delineated in Section IV. The modified framework is tested on simulated and real data, and the obtained results are discussed in Section V. Finally, remarks about the overall project and proposed future work are presented in Section VI.

II. An Overview on the Multi-Object Triangulation, Tracking and Camera Calibration Framework

The multi-object triangulation, tracking and camera calibration framework proposed by Houssineau *et al.* [12] combines the use of disparity space for modelling the uncertainty and a hierarchical structure of parent and daughter processes. The former uses a particle filter to estimate the parameters of one camera, since the other one is assumed to be the reference system. The latter estimates the state of the objects in the scene conditioned on the camera states provided by each particle in the parent process.

A general overview of the framework is presented in the following sections.

A. Representing uncertainty in the disparity space

The framework proposed in [12] is lined up with the Bayesian filtering formulation and, thus, quantification of uncertainty is required. Depending on the probability distribution exhibited on the uncertainty, different approaches could be considered. For instance, if the uncertainty follows a Gaussian distribution, the Kalman filter (KF) [16] is desired due to its simplicity; but as the system tends to be non-linear, approaches such as the Extended Kalman Filter (EKF) [17] or the Unscented Kalman filter (UKF) [18] are considered instead.

The uncertainty is commonly assumed to follow a Gaussian distribution due to the simple and yet reasonable model representation. However, as stated by the authors of the paper, a Gaussian representation is not a suitable choice in this context since the farther the object from the camera, the higher the uncertainty [19] as presented in Fig. 2. Clearly, a Gaussian distribution does not explain this behaviour accurately. Also, direct estimation of the 3D pose of an object in the Euclidean space is usually subjected to non-linear perspective projection and, hence, linearity is difficult to maintain in the process. The solution is to consider a disparity space since (i) the resulting projections from this space onto the two image planes is achieved by linear transformation, (ii) the noise on the estimations depends on the distance from the camera, and (iii) the range is actually limited by the dimensions of the image. Note that the importance of the first fact is that if the uncertainty on the state of the object of interest is represented through a Gaussian in the disparity space, the relation is kept after transforming the state back into the Euclidean space.

Under this perspective, the state of an object could be modelled by a Gaussian distribution $p_t \sim \mathcal{N}(y_t, Q_t)$ with mean y_t and covariance Q_t at time-step t in the disparity space $\mathbb{D}$. Then, the dynamics of an object of interest are described through the prediction and the update steps as follows. Initially, a particle is sampled from p_t. Then, the representation of this particle is mapped into the Euclidean space $\mathbb{X}$. Afterwards, the particle is moved in this space using the Markov transition $M_{t+1|t}$. From this point, the particle is mapped back into the disparity space. After, the obtained particle is used for predicting the Gaussian distribution $p_{t+1|t}$. Finally, the particle p_{t+1} in the next time-step $t+1$ is obtained by applying the Kalman update to $p_{t+1|t}$.

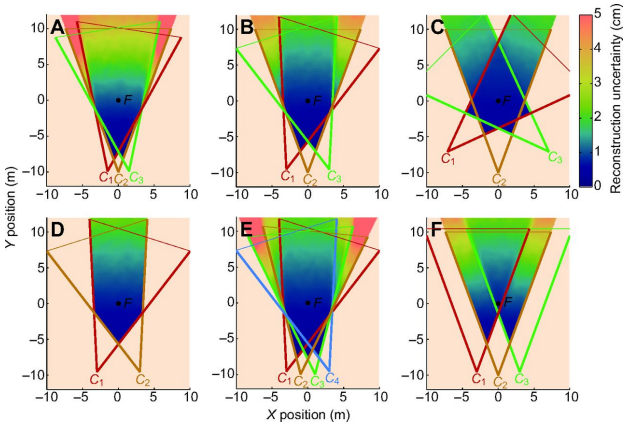

Fig. 2. Uncertainty quantification on triangulation from cameras with different setups. The farther the object to the camera, the higher the uncertainty [19].

B. Single-object estimation using non-rectified cameras

The framework presented by Houssineau *et al.*[12] is able to cope with object estimation using a non-rectified camera setup by creating artificial rectified pairs for each camera as illustrated in Fig. 3. As a result, two disparity spaces $\mathbb{D}_l$ and $\mathbb{D}_r$ are created for the left and right setup, respectively. Accordingly, the previously discussed prediction step for single-object estimation needs to be extended to handle the cases in which the observations come from one camera or the other. In the simplest case, the prediction is performed in the same camera in which the observations have been made and, hence, a usual prediction is considered. However, when the prediction is performed based on the observations of the other camera, the situation is different since the relation between the two disparity spaces should be taken into account. A procedure called *particle move* is applied, which consists in predicting in the disparity space related to the observations and then projecting this information into the other disparity space.

C. A general solution for multi-object filtering

The multi-object filtering problem consists in tracking a set of targets given that (i) the cardinality of the set is unknown and expected to vary in time since the objects of interest may disappear from the FoV or enter the scene at a certain time-step and (ii) the measurements coming from different sensors have some detection uncertainty and may represent false alarms. Thus, one of the main issues is to find the correspondence between the measurements and the targets so that the update step is performed correctly [20].

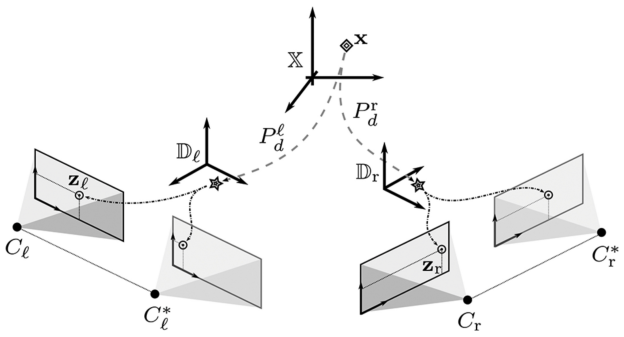

Fig. 3. Non-rectified two-cameras setup (C_l, C_r) with their corresponding abstract rectified camera pair C_l^* and C_r^*, respectively [12].

A well-known methodology for describing the multi-object filtering problem is called the Finite Set Statistics (FISST) [21]. Let $\mathcal{X}$ and $\mathcal{Z}$ be the target space and the observation space, respectively. The FISST framework encompasses the following models:

1) the multi-target state and the multi-target observation at a time-step t are defined as $X_t \in \mathcal{F}(\mathcal{X})$ and $Z_t^i \in \mathcal{F}(\mathcal{Z})$, being $\mathcal{F}(\mathcal{X})$ and $\mathcal{F}(\mathcal{Z})$ the σ-algebra on $\mathcal{X}$ and $\mathcal{Z}$, respectively;
2) at a given time-step t, each $\mathbf{x}_t \in X_t$ represented by the state $\mathbf{y}' \in \hat{\mathbb{D}}_i$ continues existing with probability $p_S(\mathbf{y}')$ or dies with a probability $1 - p_S(\mathbf{y}')$;
3) given that the object $\mathbf{x}_t \in X_t$ represented by the state $\mathbf{y}' \in \hat{\mathbb{D}}_j$ survives at time-step $t+1$, the transition from $\mathbf{x}_t$ to $\mathbf{x}_{t+1}$ with state $\mathbf{y} \in \hat{\mathbb{D}}_i$ is given by the multi-target transition density $\mathbf{M}_{t+1}^i|_t^j(\mathbf{y}|\mathbf{y}')$;
4) the multi-target state X_t at time-step t may contain a new set of objects $\mu_t^b(\mathbf{y})$ appearing spontaneously;
5) an object $\mathbf{x}_t \in X_t$ with state $\mathbf{y}' \in \hat{\mathbb{D}}_i$ is detected with probability $p_D^i(\mathbf{y}')$ or missed with probability $1 - p_D^i(\mathbf{y}')$;
6) the probability that the observation $\mathbf{z}_t$ comes from the object $\mathbf{x}_t \in X_t$ with state $\mathbf{y}' \in \hat{\mathbb{D}}_i$ is given by the multi-target likelihood $\mathbf{L}_t^i(\mathbf{z}_t|\mathbf{y}')$;
7) the observations Z_t^i at a certain time-step t are independent and may contain a set of false alarms spatially distributed by a probability distribution c whose cardinality follows a Poisson distribution with parameter λ_i.

Taking into account the previous information and accordingly to [20], [21], the estimation of the population is carried out by the prediction and update steps as follows:

$$\hat{P}_t^i(Y) = \int \mathbf{M}_t^i|_{t-1}^j(Y|Y')P_{t-1}^j(Y')\delta Y', \tag{1}$$

$$P_t^i(Y) = \frac{\mathbf{L}_t^i(Z_t|Y)\hat{P}_t^i(Y)}{\int \mathbf{L}_t^i(Z_t|Y')\hat{P}_t^i(Y')\delta Y'}. \tag{2}$$

It is important to remark that the multi-target likelihood in Eq. 2 depends on the physics of the sensors. For instance, the multi-target likelihood $L_t(\cdot|\cdot)$ may assume spurious observations whose cardinality follows a Poisson distribution.

The standard Bayes filter was initially considered for addressing this problem, but approximations are preferred instead due to the intractability of the scenario when dealing with several targets [20], [21].

D. A particular derivation from the FISST framework: the PHD filter

As we described previously, the Bayes filter computation increases exponentially as the number of targets grows. Thus, approximations of the multi-object distribution are adopted in order to ease the estimation process. These approximations are derived from the FISST framework through the first-order statistical moment of the posterior multi-target state, also known as intensity or PHD. One of the best-known derivations corresponds to the PHD filter [13], [14], [21].

For a Random Finite Set (RFS) X_t on the target space $\mathcal{X}$ with a probability distribution $\mathbb{P}$, the integral of the intensity ν on a region $B \subseteq \mathcal{X}$, i.e.

$$\int |X_t \cap B|\mathbb{P}(dX_t) = \int_B v(x_t)dx_t, \tag{3}$$

corresponds to the expected number of targets in that region [14]. Note that when $B = \mathcal{X}$, the value of the integral is equal to the total mass $\tilde{N}_t$, i.e. the expected number of targets in the whole target space.

The RFS X_t is assumed to follow a Poisson distribution [14] in the particular case of the PHD filter and, hence, the following two conditions hold:

1) each element $x_t \in X_t$ is independent from each other;
2) the elements in X_t are identically and independently distributed (i.i.d.) with probability distribution $v(\cdot)/\tilde{N}_t$.

Then, the multi-object densities $\hat{P}_t^i$ and P_t^i are propagated through their first moment densities $\hat{\mu}_t^i$ and μ_t^i as follows:

$$\hat{\mu}_t^i(\mathbf{y}) = \mu_t^b(\mathbf{y}) + \int p_S(\mathbf{y}')M_t^i|_{t-1}^j(\mathbf{y}|\mathbf{y}')\mu_t^i(\mathbf{y}')d\mathbf{y}', \tag{4}$$

$$\mu_t^i(\mathbf{y}) = (1 - p_D^i(\mathbf{y}))\hat{\mu}_t^i(\mathbf{y}) + \sum_{\mathbf{z}\in Z_t^i} \frac{p_D^i(\mathbf{y})L_t^i(\mathbf{z}|\mathbf{y})\hat{\mu}_t^i(\mathbf{y})}{\lambda^i c^i(\mathbf{z}) + \int p_D^i(\mathbf{y}')L_t^i(\mathbf{z}|\mathbf{y}')\hat{\mu}_t^i(\mathbf{y}')d\mathbf{y}'}. \tag{5}$$

In the particular approach presented by Houssineau *et al.* [12], the multi-object transition density $M_t^i|_{t-1}^j(\cdot|\cdot)$ corresponds to the *particle move* between the disparity space $\hat{\mathbb{D}}_i$ and $\hat{\mathbb{D}}_j$ from the time-step $t-1$ to t.

There are two implementations of the PHD filter: the Gaussian Mixture PHD filter [14] (GM-PHD) and the sequential Monte Carlo PHD filter [22] (SMC-PHD). Having in mind that we are using the disparity space representation presented in Section II-A and the single-object estimation approach described in Section II-B, the GM-PHD becomes a suitable solution since the Kalman filter is used for the single-object filtering and the uncertainty is represented as a Gaussian.

For using GM-PHD, the p_S and p_D are assumed to be state-independent [14]. However, constraining p_D in such way requires the use of at least three cameras [12] and, thus, this last assumption is relaxed.

The update step in implementation of the PHD filter through the GM-PHD is as follows:

$$\mu_t^i(y) = \hat{\mu}_{\circ,t}^i(y) + \sum_{z\in Z_t^i} \frac{L_t^i(z|y)\hat{\mu}_{\bullet,t}^i(y)}{\lambda^i c^i(z) + \int_{\mathcal{X}} \hat{\mu}_{\bullet,t}^i(y')dy'}, \tag{6}$$

where $\hat{\mu}^i_{\circ,t}(\cdot)$ and $\hat{\mu}^i_{\bullet,t}(\cdot)$ represent the missed detection and associated terms, respectively. Both terms can be expressed as follows:

$$\hat{\mu}^i_{\circ,t}(y) = \sum_{k=1}^{N}(1 - p^i_D(y))\omega_k \mathcal{N}(y, \hat{y}^i_k, \hat{Q}^i_k), \tag{7}$$

$$\hat{\mu}^i_{\bullet,t}(y) = \sum_{k=1}^{N} p^i_D(y)\omega_k \mathcal{N}(y, \hat{y}^i_k, \hat{Q}^i_k). \tag{8}$$

E. The parent process: the particle filter

We have seen how to perform multi-object filtering if the states of the camera are given in advance. However, their extrinsic and intrinsic parameters might, in fact, not be provided as inputs and, thus, the idea behind the parent process of the framework is to determine them along with the multi-object estimation. This task is difficult to address due to the number of intrinsic and extrinsic parameters to be determined in the problem. One way to deal with this problem, for instance, is to assume that the world coordinate system is the one of the left camera and, hence, the framework requires calculating up to 16 values for the state of the right camera.

Formally, the problem consists in finding the joint probability distribution $\mathbf{P}^i_t$ describing the state of the right camera $\mathbf{s} \in \mathbb{S}_r$ as well as the state of the objects of interest Y [12]:

$$\mathbf{P}^i_t(Y, \mathbf{s}) = P^i_t(Y|\mathbf{s})p_t(\mathbf{s}), \tag{9}$$

where $p_t(\cdot)$ is the probability distribution over $\mathbb{S}_r$ at time-step t.

As we have discussed in previous sections, the multi-object density is computationally hard to process directly and, thus, its first-order moment statistic is used. The derivation of the first-order moment density differs from the usual PHD filter expressions since, now, the multi-object density is conditioned on the state of the right camera. The corresponding expression following the arguments in [23] is

$$\mu^i_t(\mathbf{y}, \mathbf{s}) = \mu^i_t(\mathbf{y}|\mathbf{s})\alpha_t(\mathbf{s})p_t(\mathbf{s}), \tag{10}$$

where $\mu^i_t(Y|\mathbf{s})$ comes from the daughter process conditioned on s and $\alpha_t(\mathbf{s}) \in [0, 1]$, expressed as

$$\alpha_t(\mathbf{s}) = \frac{\mathbf{L}^c_t(Z^i_t|\mathbf{s})}{\int \mathbf{L}^c_t(Z^i_t|\mathbf{s}')p_t(\mathbf{s}')d\mathbf{s}'}, \tag{11}$$

represents the probability that $\mathbf{s}$ generates correct multi-object update. In the previous expression, $\mathbf{L}^i_t(Z^i_t|\mathbf{s})$ is the likelihood of the observations given the state $\mathbf{s}$, defined as below:

$$\mathbf{L}^c_t(Z^i_t|\mathbf{s}) = \exp\left(-\lambda(\mathbf{s}) - \int \mu^i_{\bullet,t}(\mathbf{y}|\mathbf{s})d\mathbf{y}\right) \cdot \prod_{z \in Z^i_t}\left[\lambda(\mathbf{s})c(\mathbf{z}|\mathbf{s}) + \int L^i_t(\mathbf{z}|\mathbf{y}, \mathbf{s})\mu^i_{\bullet,t}(\mathbf{y}|\mathbf{s})d\mathbf{y}\right]. \tag{12}$$

If the probability distribution p_t is approached through a set of M_t particles $\{s_k\}_{k=1}^{M_t}$, i.e.

$$p_t(\mathbf{s}) = \sum_{k=1}^{M_t}\omega_k \delta_{\mathbf{s}_k}(\mathbf{s}), \tag{13}$$

$$\delta_{\mathbf{s}_k}(\mathbf{s}) = \begin{cases} 1 & \text{if } \mathbf{s} = s_k \\ 0 & \text{otherwise,} \end{cases} \tag{14}$$

then, the expression for the first-moment density μ^i_t is approximated as follows:

$$\mu^i_t(\mathbf{y}, \mathbf{s}) \approx \sum_{k=1}^{M_t}\mu^i_t(\mathbf{y}|\mathbf{s}_k)\alpha_t(\mathbf{s}_k)\omega_k\delta_{\mathbf{s}_k}(\mathbf{s}). \tag{15}$$

Therefore, each particle of the parent process is propagated using the conditioned GM-PHD filter of the daughter process.

III. Introducing variance on the framework

The variance in the number of targets observed in the scene can give information regarding the reliability of the measurements coming from the sensing system [15]. This knowledge could be strategically used for determining whether the current configuration of the right camera is correct or not.

In the following sections, we described how the variance is plugged into the framework, i.e. the parent and the daughter processes.

A. Variance in the parent process

The overall variance in the number of targets in the scene could be calculated based on different approaches. For instance, in this particular case, we consider (i) the variance information on the most likely particle, and (ii) the expectation of the variance with respect to the information of all the particles. The two approaches for addressing the variance in the framework are analysed in the experimentation section.

B. Variance in the daughter process

As we described previously, the PHD filter propagates only first-order moments. However, the variance of the updated target process of the camera C_i with the set of measurements Z^i_t can be calculated at a given time-step t by considering the regional variance approach presented in [15]. Then, the variance on a region $B \subseteq \mathcal{X}$ is expressed as

$$\begin{aligned}\mathrm{Var}^i_t(B|Z^i_t) = &\int_B \hat{\mu}^i_{\circ,t}(\mathbf{y})d\mathbf{y} \\ &+ \sum_{\mathbf{z}\in Z^i_t}\frac{\int_B L^i_t(\mathbf{z}|\mathbf{y})\hat{\mu}^i_{\bullet,t}(\mathbf{y})d\mathbf{y}}{\lambda^i c^i(\mathbf{z}) + \int_{\mathcal{X}} L^i_t(\mathbf{z}|\mathbf{y})\hat{\mu}^i_{\bullet,t}(\mathbf{y})d\mathbf{y}} \\ &\cdot\left(1 - \frac{\int_B L^i_t(\mathbf{z}|\mathbf{y})\hat{\mu}^i_{\bullet,t}(\mathbf{y})d\mathbf{y}}{\lambda^i c^i(\mathbf{z}) + \int_{\mathcal{X}} L^i_t(\mathbf{z}|\mathbf{y})\hat{\mu}^i_{\bullet,t}(\mathbf{y})d\mathbf{y}}\right).\end{aligned} \tag{16}$$

In this particular case, we are not interested in computing the variance on a specific region, but on the whole state space $\mathcal{X}$.

IV. Implementation

In the previous sections, we describe how multi-object filtering works with the GM-PHD filter. In this section, we describe the strategy for implementing the data update and obtaining statistics information. The pseudo-code is presented in Algorithm 1. Note that we followed a similar notation to the one used in [14].

Algorithm 1: Data update on the PHD filter and information statistics

Data: Mixture information $\{\omega_k, \hat{y}_k^i, \hat{Q}_k^i\}_{k=1}^N$, states $Y = \{y_1, y_2, ..., y_N\}$ and new measurements Z_t^i
Result: Expectation and variance in number of targets $\mu(\mathcal{X})$ and $\mathrm{Var}(\mathcal{X}|Z_t^i)$, respectively

```
/* Missed and associated terms
   computation                                   */
```
for $k = 1$ ***to*** N **do**
  // Computing missed detections
  $\omega_k^\circ \leftarrow (1 - p_D^i(y_k))\omega_k \mathcal{N}(y_k, \hat{y}_k^i, \hat{Q}_k^i)$;
  // Computing associated terms
  foreach $z_j \in Z_t^i$ **do**
    $\hat{\omega}_{k,z_j}^\bullet \leftarrow p_D^i(y_k) L_t^i(\mathbf{z_j}|\mathbf{y})\omega_k \mathcal{N}(y_k, \hat{y}_k^i, \hat{Q}_k^i)$;

```
/* Update step                                   */
```
for $k = 1$ ***to*** N **do**
  // Normalising measurement weights
  foreach $z_j \in Z_t^i$ **do**
    $\omega_{k,z_j}^\bullet \leftarrow \hat{\omega}_{k,z_j}^\bullet / \left(\sum_{k'} \hat{\omega}_{k',z_j}^\bullet + \lambda^i c^i(z_j)\right)$;
  // Updating weights
  $\omega_k \leftarrow \omega_k^\circ + \sum_{z_j \in Z_t^i} \omega_{k,z_j}^\bullet$;

```
/* Statistics calculation                        */
// Expectation in number of targets
```
$\mu(\mathcal{X}|Z_t^i) \leftarrow \sum_k \omega_k$;
```
// Variance in number of targets
```
$\omega_{z_j}^\bullet \leftarrow \left[\sum_k \hat{\omega}_{k,z_j}^\bullet\right] / \left[\sum_k \hat{\omega}_{k,z_j}^\bullet + \lambda^i c^i(z_j)\right]$;
$\mathrm{Var}(\mathcal{X}|Z_t^i) \leftarrow \sum_k \omega_k^\circ + \sum_{z_j \in Z_t^i} \omega_{z_j}^\bullet$;

V. Results on simulated and real data

The proposed approach was evaluated on a simulated scenario when considering different amounts of particles on the parent process. The obtained results were used to guide the decision of how many particles to use on the GM-PHD filter when running the trials on real data.

A. Experiments on simulated data

Generating simulated data following a specific scenario setup, i.e. the ground truth, was essential to validate the proposal. The camera pair was symmetrically located with respect to the reference frame; the left camera C_l was at $(-20, 0, 0)$ and rotated $\pi/12$ with respect to the y axis, while the right camera C_r was at $(20, 0, 0)$ and rotated $-\pi/12$ around the y axis. Then, a total of seven targets moving at constant velocity were introduced in the scene in specific time-steps: (i) 2 targets were in the scene from the beginning of the experiment, (ii) other two entered the scene at time-steps 20 and 35, respectively, (iii) other two targets were simultaneously introduced at time-step 70 and, (iv) finally, the seventh target joined the scene at time-step 120. During the experiment, three of the targets got out of the FoV of the cameras at time-steps 123, 229 and 324, in that order. The total number of targets at each time-step is indicated in Fig. 4 with a green line.

By knowing the scenario setup as well as the position of the targets in the 3D space, the corresponding projections on the image plane of each camera were computed. Before processing such information with the proposed extension of the camera calibration framework of Houssineau *et al.* [12], its parametrisation was carried out: the probability of detection p_D^i was set to 0.95, the merging distance was 7, the pruning threshold was equal to 10^{-6}, and the false alarm Poisson parameter was set to $\lambda^i = 1$. As reported in Fig. 4, not only the variance of the expected number of targets in the scene was computed with the two approaches introduced in Section II and Section III, but also with three different amount of particles: 250, 500 and 1500. Each of those experiments was executed once.

From the results presented in Fig. 4, it can be seen that the performance of the proposed approach is directly related to the existing number of particles in the parent process. When considering more particles and indifferently from the followed approach to compute the statistics, the estimated number of targets on the scene represents the reality better, i.e. the estimation is closer to the ground truth. Despite the statistics being more stable when increasing the number of particles, it leads to a slower convergence of the expectation to the real value.

Comparing the two approaches for computing the statistics, it can be clearly spotted out that considering the information of one particle gives more stable results and faster convergence than computing the weighted-average of the information of all the particles. Additionally, further experiments have demonstrated that the performance of the weighted-average approach plummets when the particles are not resampled properly.

Regarding the variance of the estimated number of targets in the scene, it has been seen that it slightly increases when (a) the number of tracked targets grows, (b) a new target appears in the scene or (c) a target goes out of the FoV of the cameras. The former fact can be spotted in any of the illustration in Fig. 4 by comparing the width of the shadowed area when a different number of targets are estimated. The two last points cause small sudden spikes in the variance, which are not noticeable in the figures due to its reduced size. A quantitative analysis of the obtained results also indicates that the variance is a bit lower when the statistics are computed with the most likely particle. However, non-consistent conclusions could be drawn regarding the behaviour of the variance when increasing the number of particles.

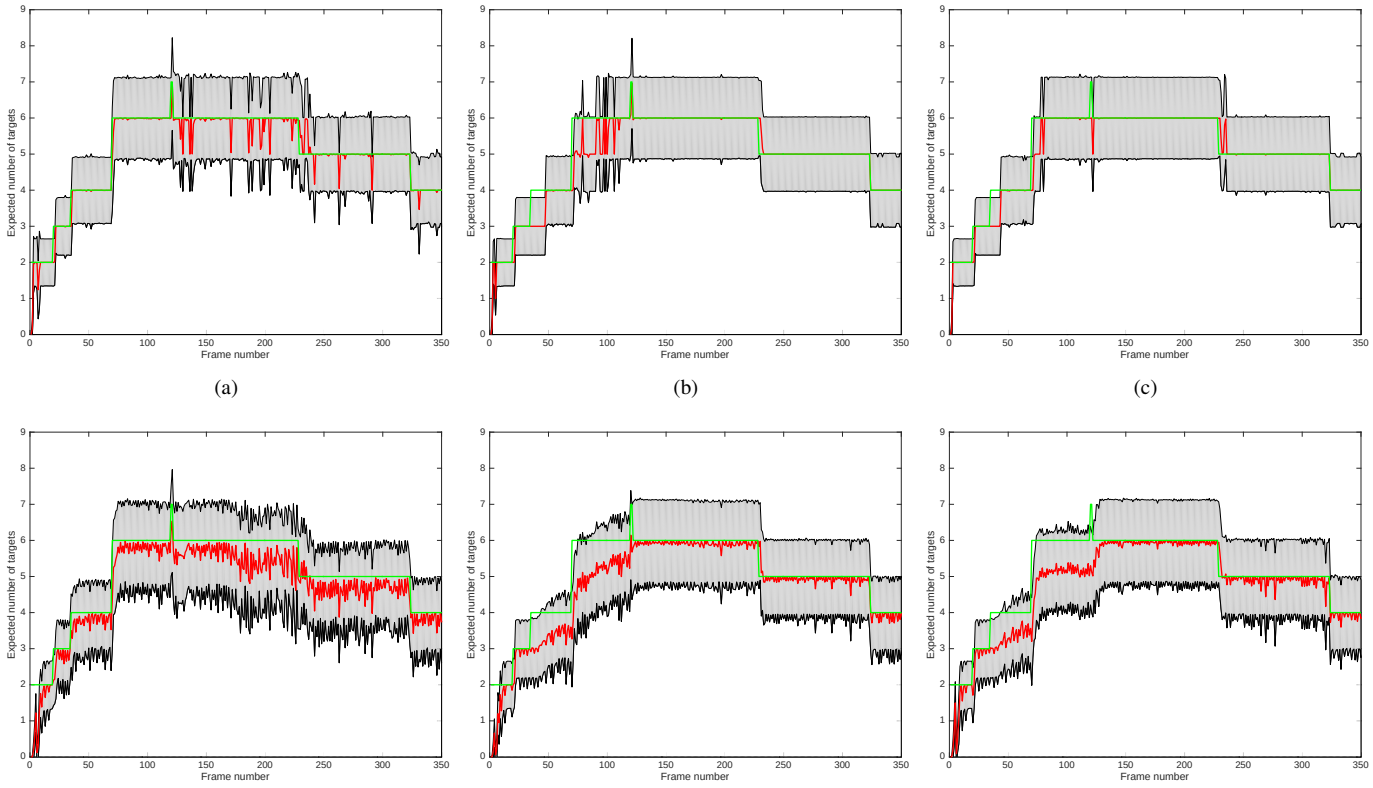

Fig. 4. Real number of targets in the scene (green line), and obtained expectation (red line) and $2-\sigma$ confidence level (shadowed area), where σ corresponds to the standard deviation in the number of targets in the scene at each time-step. First row, statistics when considering the information of the particle with the highest weight. Second row, statistics when considering the weighted average of the information of all the particles. From the column on the left to the one on the right, obtained results when using 250, 500 and 1500, respectively.

B. Experiments on real data

The evaluation of the proposal on real data was performed using the dataset gathered by Houssineau *et al* [12]. The objects of interest in this dataset are paper planes, which sustain their flight for some time after they are launched. Each of those planes is observed by the pair of cameras as shown in Fig. 5; Fig. 5a and 5b correspond to the time-step 190, where only one plane is in the scene, and Fig. 5c and 5d depict three targets at time-step 240.

For this experiment, the parametrisation of the GM-PHD is similar to the one considered in Section V-A, i.e. the probability of detection p_D^i was equal to 0.95, the merging distance was set to 7, the pruning threshold was 10^{-6} and the false alarm Poisson parameter was equal to $\lambda^i = 1$. The number of used particles was set to 1500 accordingly to the intuition given in Section V-A.

Two different approaches for computing the mean and variance in the expected number of targets in the scene have been introduced in Section II and Section III. When only the particle with the highest likelihood is considered for determining the statistics, the obtained results look like in Fig. 6. On the other hand, if the information of all the particles is weighted by its likelihood and then averaged, the obtained statistics are shown in Fig. 7. In both cases, the estimated number of targets at time-steps 190 and 250 agree with the visual ground truth provided in Fig. 5.

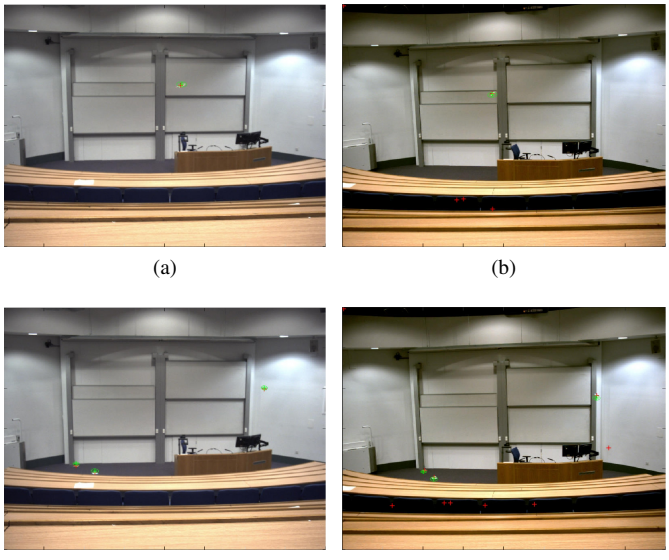

Fig. 5. Estimated target mean and variance (green crosses and ellipses, respectively) and observations (red crosses). First row: frame 190. Second row: frame 240. Left and right columns, left and right view, respectively.

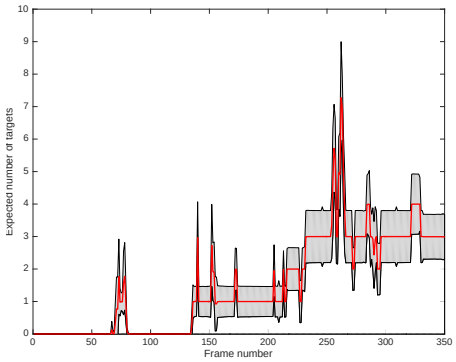

Fig. 6. Expected number of targets when considering the information of the most likely particle.

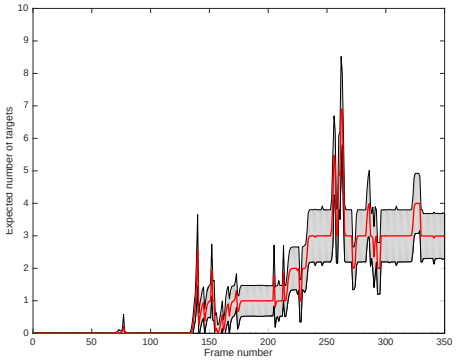

Fig. 7. Expected number of targets when considering the weighted average of the information of all the particles.

The obtained results support the conclusions stated in Section V-A; averaging the information of all the particles, instead of considering only the information of the most likely particle, increases the response time. For instance, this fact is reflected in time-step 75, where no targets are estimated when using the former method in contrast to the two estimated targets when considering the latter approach.

VI. Final remarks

In this paper, an approach for estimating the variance in the number of targets on a framework performing camera calibration from multi-object triangulation and tracking is presented.

The proposed approach was evaluated on real and simulated scenarios. It was observed that the more the number of particles in the parent process, the better the approximation of the number of targets and the slower the convergence to the expected value. The variance did not exhibit a strict behaviour regarding the number of particles, but with the number of targets on the FoV of the sensing system.

Extensions of this work will contemplate weighting the particles in the parent process not only from the current measurement model, but also from the variations in the number of targets.

Acknowledgment

The authors would like to thank the Education, Audiovisual and Culture Executive Agency (EACEA) of the European Commission for the received Erasmus Mundus MSc Course (EMMC) scholarships, which have made possible to undertake the MSc in VIsion and roBOTics (VIBOT) and thus, this work.

References

[1] R.I. Hartley and P. Sturm. Triangulation. *Computer vision and image understanding*, 68(2):146–157, 1997.

[2] N. Ayache. *Artificial vision for mobile robots: stereo vision and multisensory perception*. Mit Press, 1991.

[3] U.R. Dhond and J.K. Aggarwal. Structure from stereo-a review. *IEEE transactions on systems, man, and cybernetics*, 19(6):1489–1510, 1989.

[4] D. Scharstein and R. Szeliski. A taxonomy and evaluation of dense two-frame stereo correspondence algorithms. *International journal of computer vision*, 47(1-3):7–42, 2002.

[5] J. Davis, R. Ramamoorthi, and S. Rusinkiewicz. Spacetime stereo: a unifying framework for depth from triangulation. In *Computer Vision and Pattern Recognition, 2003. Proceedings. 2003 IEEE Computer Society Conference on*, volume 2, pages II–359–66 vol.2, June 2003.

[6] Z. Zhang. Determining the epipolar geometry and its uncertainty: A review. *International journal of computer vision*, 27(2):161–195, 1998.

[7] H. C. Longuet-Higgins. A computer algorithm for reconstructing a scene from two projections. *Readings in Computer Vision: Issues, Problems, Principles, and Paradigms, MA Fischler and O. Firschein, eds*, pages 61–62, 1987.

[8] R.Y. Tsai and T. S. Huang. Uniqueness and estimation of three-dimensional motion parameters of rigid objects with curved surfaces. *IEEE Transactions on pattern analysis and machine intelligence*, 6(1):13–27, 1984.

[9] X. Armangué and J. Salvi. Overall view regarding fundamental matrix estimation. *Image and vision computing*, 21(2):205–220, 2003.

[10] A. Fusiello, E. Trucco, and A. Verri. A compact algorithm for rectification of stereo pairs. *Machine Vision and Applications*, 12(1):16–22, 2000.

[11] 3D stereo camera calibration. http://www.ivs.auckland.ac.nz/web/calibration.php. Accessed: 2016-11-19.

[12] J. Houssineau, D.E. Clark, S. Ivekovic, C. S. Lee, and J. Franco. A unified approach for multi-object triangulation, tracking and camera calibration. *IEEE Transactions on Signal Processing*, 64(11):2934–2948, June 2016.

[13] R. Mahler. Multitarget bayes filtering via first-order multitarget moments. *IEEE Transactions on Aerospace and Electronic Systems*, 39(4):1152–1178, Oct 2003.

[14] B-N. Vo and W-K Ma. The gaussian mixture probability hypothesis density filter. *IEEE Transactions on signal processing*, 54(11):4091–4104, 2006.

[15] E. Delande, M. Üney, J. Houssineau, and D. E. Clark. Regional variance for multi-object filtering. *IEEE Transactions on Signal Processing*, 62(13):3415–3428, July 2014.

[16] E. R. Kalman. A new approach to linear filtering and prediction problems. *Journal of basic Engineering*, 82(1):35–45, 1960.

[17] A. H. Jazwinski. *Stochastic processes and filtering theory*. Courier Corporation, 2007.

[18] E. A. Wan and R. Van Der Merwe. The unscented Kalman filter for nonlinear estimation. In *Adaptive Systems for Signal Processing, Communications, and Control Symposium 2000. AS-SPCC. The IEEE 2000*, pages 153–158. Ieee, 2000.

[19] D. H. Theriault, N. W. Fuller, B. E. Jackson, E. Bluhm, D. Evangelista, Z. Wu, M. Betke, and T.L. Hedrick. A protocol and calibration method for accurate multi-camera field videography. *Journal of Experimental Biology*, pages jeb–100529, 2014.

[20] R. Mahler. PHD filters of higher order in target number. *IEEE Transactions on Aerospace and Electronic Systems*, 43(4):1523–1543, October 2007.

[21] B. T. Vo. *Random finite sets in multi-object filtering*. Citeseer, 2008.

[22] B-N Vo, S. Singh, and A. Doucet. Sequential monte carlo methods for multitarget filtering with random finite sets. *IEEE Transactions on Aerospace and electronic systems*, 41(4):1224–1245, 2005.

[23] B. Ristic, D. E. Clark, and N. Gordon. Calibration of multi-target tracking algorithms using non-cooperative targets. *IEEE Journal of Selected Topics in Signal Processing*, 7(3):390–398, June 2013.

Èric Pairet received the B.E. in Electronics and Automation Engineering in 2015 from the University of Girona, Girona, Spain. He is a member of the Research Center in Underwater Robotics (CIRS) at the same university and he is currently enrolled in an Erasmus Mundus Master in Computer Vision and Robotics (VIBOT) at the Universities of Burgundy (France), Girona (Spain) and Heriot-Watt (UK). His research interests include: intelligent control architectures, uncertainty in systems, machine learning and computer vision.

Jose Bernal received the B.E. in Computer Engineering in 2014 from the Universidad del Valle, Cali, Colombia. Currently, he is enrolled in an Erasmus Mundus Master in Computer Vision and Robotics (VIBOT) at University of Burgundy, University of Girona and Heriot-Watt University. His research interests include: computer vision, medical imaging, segmentation techniques, numerical methods and optimisation techniques.

Isabel Schlangen is a current Ph.D. studentship holder at the Edinburgh Super-Resolution Imaging Consortium, UK. She received a German diploma in mathematics from the University of Bonn (Germany) in 2012 and a joint M.Sc. degree in vision and robotics from the Universities of Burgundy (France), Girona (Spain), and Heriot-Watt (UK) in 2014. Her current research interests are multi-target estimation, probability theory and image analysis in a mainly biomedical context.

Daniel E. Clark is an Associate Professor in the School of Engineering and Physical Sciences at Heriot-Watt University. His research interests are in the development of the theory and applications of multi-object estimation algorithms for sensor fusion problems. He has led a range of projects spanning theoretical algorithm development to practical deployment. He was awarded his Ph.D. in 2006 from Heriot-Watt University.