

# EFFICIENT ESTIMATION OF TARGET DETECTION QUALITY

Juan C. SanMiguel<sup>1</sup> and Andrea Cavallaro<sup>2</sup>

<sup>1</sup>Universidad Autónoma de Madrid (Spain), <sup>2</sup>Queen Mary University of London (UK)

## ABSTRACT

The capability of determining the quality of target detections is important for applications using smart cameras, such as autonomous robotics and surveillance. We propose to estimate the quality of target detections by integrating the target location uncertainty over polygonal domains, which represent the fields of view of the cameras. We define a framework based on numerical integration that easily accommodates multiple models for uncertainty and fields of view. We perform quadrature-based integration combined with importance sampling to provide accurate quality estimations while reducing the computational cost. The proposed method outperforms alternative approaches in terms of estimation accuracy and execution time. We validate the proposed approach with a recent distributed multi-camera multi-target tracker and improved it by considering realistic fields of view. Results demonstrate the effectiveness of the proposed method in decreasing the state estimation error.

**Index Terms**— Detection quality, quadrature-based integration, multi-target tracking, camera networks.

## 1. INTRODUCTION

Detection quality is related to the probability of a target to be detected within the Field Of View (FOV) of a camera [1]. This detection probability models the miss-detection rate and accounts, over time, for the number of undetected targets that are within the FOV. Tracking algorithms rely on accurate estimations of detection quality. For example, Probabilistic Data Association [2] and Markov Chain Monte Carlo Data Association [3] employ quality estimations to compute the weights for associating (new) observations to (existing) targets. The observation likelihood is generally used as proxy for detection quality to measure the distance between a target model and a detection hypothesis [4].

Parametric quality prediction models acquired from training data can be used to predict performance [5]. Detailed target modeling may help applications such as multi-view face detection and [6] and 3D object recognition [7]. However, these methods have limited generalization to other tasks of camera networks.

The probability of (correctly) detecting a target within the camera FOV is often set as a constant, whose value reflects prior knowledge on the miss-detection rate [8, 9]. FOV models can generalize the sensing capabilities in a heterogeneous camera network composed for example of zenithal aerial views and perspective surveillance cameras [10]. These models describe the geometrical properties of the FOV whose projection onto a 2D common plane is a widely accepted FOV abstraction [11]. These projections take the shape of quadrilaterals or triangles. To calculate the detection probability, squared FOVs enable fast computation via the cumulative distribution function of the predicted target location [12]. However,

triangular FOVs are more common in real applications [11], but do not lend themselves to the same fast computation strategy.

To address this problem, we propose a generic estimator for target detection quality. We map the problem to a definite integral where the integration domain represents the FOV of a camera. This integral is numerically approximated by combining quadrature-based integration and importance sampling. Moreover, we empirically derive a closed-form function to determine the optimal number of integration points, based on system uncertainty models. As validation, we use the proposed method in the data association stage of a recent distributed multi-camera multi-target tracker [12]. Results show the effectiveness of the proposed estimator both in terms of tracking performance and of accuracy-complexity trade-off of the integration.

The rest of the paper is structured as follows. Section 2 discusses the elements contributing to target detection quality. Section 3 describes the proposed estimator. Section 4 presents the experimental results and the application to multi-target tracking. Finally, Section 5 concludes the paper.

## 2. PROBABILITY OF TARGET DETECTION

Let  $\{t_j\}_{j=1}^{N_T}$  be a set of  $N_T$  moving targets and  $\mathbf{x}_j(k)$  be the state of each target at time  $k$ . The motion model for target  $j$  is

$$\mathbf{x}_j(k+1) = \Phi \mathbf{x}_j(k) + \Gamma_j(k), \quad (1)$$

where  $\Phi \in \mathcal{R}^{4 \times 4}$  is the state transition matrix and  $\Gamma_j(k)$  is the process noise, which is modeled as  $\mathcal{N}(0, Q)$ , where  $Q$  is the process noise covariance matrix.

Let  $\{C_i\}_{i=1}^{N_C}$  be a network of  $N_C$  cameras that observe the targets. Each camera  $i$  generates a set  $\{\mathbf{z}_i^n(k)\}_{n=0}^{N_O^i(k)}$  of  $N_O^i(k)$  observations as follows:

$$\mathbf{z}_i^n(k) = H_i^j \mathbf{x}_j(k) + v_i^j(k), \quad (2)$$

where  $H_i^j \in \mathcal{R}^{2 \times 4}$  is the measurement matrix that maps the target states to the observation state space [12] and  $v_i^j(k)$  is the measurement noise, which is modeled as  $\mathcal{N}(0, R)$ , where  $R$  is the measurement noise covariance matrix.

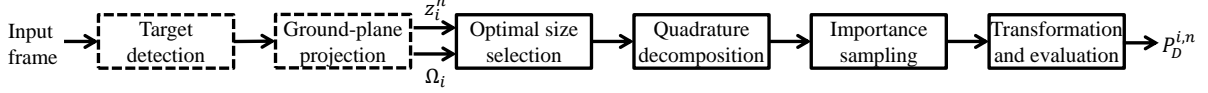
An observation  $\mathbf{z}_i^n(k)$  can be obtained with pedestrian detectors [8, 4] and can be represented as a point [12, 13] or a bounding box.<sup>1</sup> Adapting the formulation from [14], the probability  $P_D^{i,n}$  to detect target  $n$  in camera  $i$  is

$$P_D^{i,n} = \int_{\Omega_i} P(\text{vis}/\mathbf{z}_i^n, \mathbf{x}_j^-) f(\mathbf{z}_i^n) d\mathbf{z}, \quad (3)$$

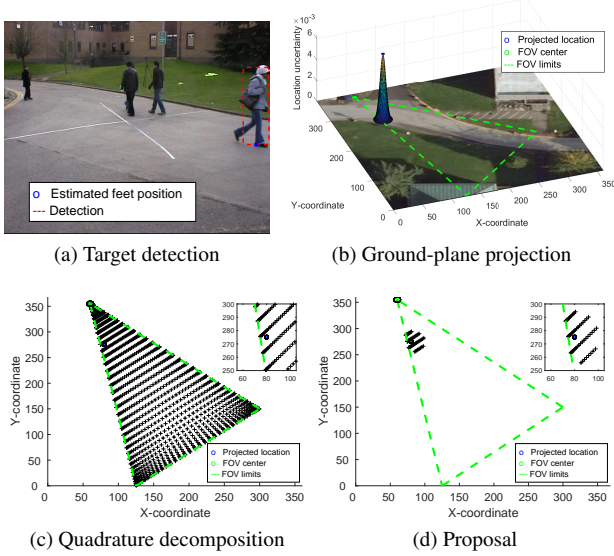
where  $f(\mathbf{z}_i^n)$  determines the detection uncertainty, which is usually modeled as a Gaussian distribution [14];  $\Omega_i$  is a polygon that defines

J. C. SanMiguel acknowledges the support of the Spanish Government through the HAVIDEO project under Grant TEC2014-5317-R.

<sup>1</sup>To improve readability we drop the time index  $k$  in the rest of the paper.



**Fig. 1.** Block diagram of the proposed approach to estimate target detection quality. The blocks with dashed lines provide the input to the estimation, namely a detection  $z_i^n$ , the camera FOV  $\Omega_i$  and the associated uncertainty model  $f(z_i^n)$ .



**Fig. 2.** Illustration of the main steps of the proposed approach.

the FOV of camera  $i$ ; and  $P(\text{vis}/z_i^n, \mathbf{x}_j^-)$  is a binary indicator for the observability (1) or non-observability (0) of the target based on the predicted target state  $\mathbf{x}_j^-$ . Such observability should consider that the target of interest may be occluded (by static scene elements and other targets) or may be outside the camera FOV.

Our goal is to accurately and quickly compute  $P_D^{i,n}$  for the camera FOV  $\Omega_i$ .

### 3. DETECTION QUALITY ESTIMATION

We cast the problem of detection quality estimation as the definite integral of  $f(\mathbf{z}_i^n)$  over the FOV domain  $\Omega_i$ . We estimate the integral in four steps (see Fig. 1 and Fig. 2), which are described below.

We numerically approximate the value of the integral for  $f(\mathbf{z}_i^n)$  using quadrature-based rules [15]

$$\int_{\Omega_i} f(\mathbf{z}_i^n) dz = \sum_{r=1}^N w_r f(\hat{\mathbf{z}}_r), \quad (4)$$

where  $N$  is the number of selected samples,  $\hat{\mathbf{z}}_r$  are the selected points in the 2D domain defined by  $\Omega_i$  and  $w_r$  are weights. To compute each  $w_r$  we use the Gauss-Legendre quadrature:

$$w_r = \frac{2}{(1 - (\hat{\mathbf{z}}_r)^2) [P'_N((\hat{\mathbf{z}}_r))]^2}, \quad (5)$$

where  $P'_N(\cdot)$  are the Legendre polynomials [15].

Assuming Gaussian probability distribution functions (pdf) for  $f(\mathbf{z}_i^n)$ , increasing the number of samples leads to a decrease of the

integration error, up to a limit [16]. However, to evaluate these additional samples the associated cost increases. Finding an optimal accuracy-cost compromise is key for an efficient estimation.

Adjusting the FOV may be needed to maximize video analytics performance. For example, some FOVs may be narrowed to acquire high-resolution views of targets whereas other FOVs may be widened to get an overall perspective [17]. We are therefore interested in the relationship between the area covered by the FOV of camera  $i$  and the spatial uncertainty associated to each  $\mathbf{z}_i^n$ .

FOV adaptations change the FOV area on the ground plane, thus requiring additional samples to maintain the integration accuracy when larger 2D areas need to be covered. Moreover, decreasing the uncertainty about the location of an observation requires an increase of the number of integration points to efficiently sample the integrand  $f(\mathbf{z}_i^n)$ .

We obtain the number of samples  $N^*$  as:

$$N^* = \underset{N}{\operatorname{argmin}} h(N, FOV_i^a, \Sigma), \quad (6)$$

where  $h(\cdot)$  defines the relation between the FOV area  $FOV_i^a$ , the covariance matrix  $\Sigma$  determining the location uncertainty of  $\mathbf{z}_i^n$ , and the number of points employed  $N$ . Empirically, we found the following relationship:

$$h(N, FOV_i^a, \Sigma) = \left| N - \frac{FOV_i^a}{\Sigma} \right|, \quad (7)$$

where  $|\cdot|$  is the L1-norm.

We use a higher density of points around the location where the integrand  $f(\mathbf{z}_i^n)$  exhibits high values. Such location corresponds to the expectation of the pdf describing  $f(\mathbf{z}_i^n)$  which is often assumed Gaussian [14] so the location corresponds to  $\mathbf{z}_i^n$  (see Eq. 3 and Eq. 2).

We re-express the integral in Eq. 4 to consider that this integrand is non-zero only in specific regions of the domain  $\Omega_i$ :

$$I = \int_{\Omega_i} \frac{f(\mathbf{z}_i^n)}{g(\mathbf{z}_i^n)} g(\mathbf{z}_i^n) dx, \quad (8)$$

where  $g(\cdot)$  is an auxiliary function to indicate which regions are significant (i.e.  $f(\mathbf{z}_i^n)$  is large) in the camera FOV  $\Omega_i$ .

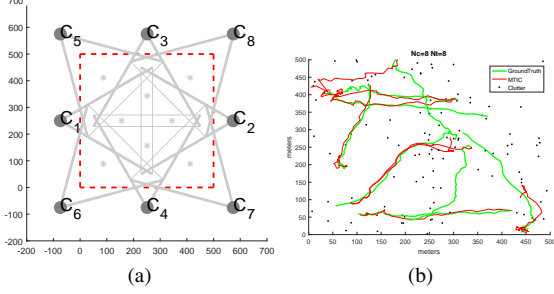
Standard importance sampling draws a set  $\{\tilde{\mathbf{z}}_1, \dots, \tilde{\mathbf{z}}_N\}$  of  $N$  samples from  $g(\cdot)$  and approximates the integral as

$$\bar{I} = \frac{1}{N} \sum_{r=1}^N \frac{f(\tilde{\mathbf{z}}_r)}{g(\tilde{\mathbf{z}}_r)}. \quad (9)$$

We define  $g(\cdot)$  as a gating region determined by the uncertainty of the observation location and then extract the selected  $N$  samples using the previously described quadrature decomposition<sup>2</sup>.

The Gaussian quadrature for triangular domains may employ accurate methods specifically for triangles, such as the Dunavant rules [18]. However, such rules are not defined for any given number of

<sup>2</sup>Alternatively, the set of samples can be generated with Monte Carlo sampling methods [16].



**Fig. 3.** Multi-camera multi-target setup: (a) the camera network and (b) clutter observations affecting the tracking process and sample trajectories.

samples (e.g. Dunavant rules are up to 20 samples), which limits their use for different FOV sizes where variable number of samples may be needed. Another option is to generate an initial sample set within a square domain (which is easier to get than triangles) and then compute the desired samples by collapsing the square to a unit triangle [19]. Although suboptimal, this square-to-triangle mapping allows one to generate any number of samples to consider variable-size FOVs.

We tabulate  $\hat{z}_r$  and  $w_r$  for various values of  $N$  in a unitary triangle. Then we select the optimal value of  $N$  for each  $\Omega_i$  and  $\Sigma$  and we obtain the tabulated samples from the unitary triangle, which are filtered according to importance sampling. Finally, we convert the coordinates of the selected samples to the desired FOV to integrate  $f(\mathbf{z}_i^n)$ .

From this unitary triangle we use an affine transformation  $T$  to any triangular shape that represents the camera FOV. The transformation  $T$ , defined as  $T = X \cdot A^{-1}$ , has 6 Degrees of Freedom (DoF) and can be fully developed as

$$\begin{bmatrix} t_1 & t_2 & t_3 \\ t_4 & t_5 & t_6 \\ 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} x_1 & y_1 & z_1 \\ x_2 & y_2 & z_2 \\ 1 & 1 & 1 \end{bmatrix} \cdot \begin{bmatrix} a_1 & b_1 & c_1 \\ a_2 & b_2 & c_2 \\ 1 & 1 & 1 \end{bmatrix}^{-1} \quad (10)$$

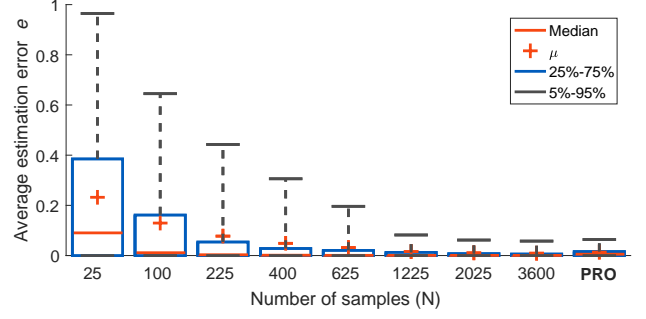
The samples in the desired triangular shape are obtained as  $X = T \cdot A$ . Finally, we obtain  $P_D^{i,n}$  for each observation  $\mathbf{z}_i^n$  in camera  $i$  as the weighted average of the transformed samples with  $T$  (Eq. 4) multiplied by the total area of the FOV  $\Omega_i$ .

#### 4. APPLICATION TO DISTRIBUTED MULTI-CAMERA MULTI-TARGET TRACKING

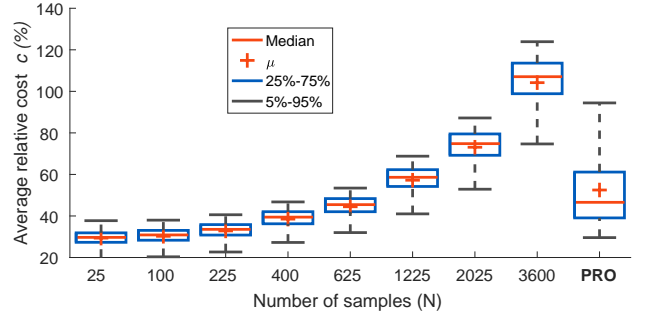
The MTIC multi-target tracker [12] uses square FOVs and computes  $P_D$  by integrating the *pdf* of the predicted innovations (i.e. the difference between  $\mathbf{z}_i^n$  and the prediction  $\mathbf{x}_j^-$ ).  $P_D$  is used for target-observation association by Probabilistic Data Association filtering (PDA). Then, the cameras exchange their estimations to achieve a consensus on each target state.

We improve the model of the MTIC multi-target tracker [12] and apply the proposed approach. Our improvement to MTIC is twofold. First, we consider triangular FOVs, which are more widely used with 5 DoF: 2D position, orientation, depth and viewing angle. Second, we use the proposed approach for fast computation of  $P_D$  in triangular FOVs.

For validation we consider a network of  $N_C = 8$  cameras [13] (see Fig. 3a) with partially overlapping FOVs  $\Omega_i$  modeled as triangles with 400m range and  $45^\circ$  angle. The communication capa-



(a) Overall accuracy



(b) Relative cost

**Fig. 4.** Comparison of detection quality estimation for the proposed approach (PRO) and the quadrature-based integration with varying number of samples (Eq. 4).

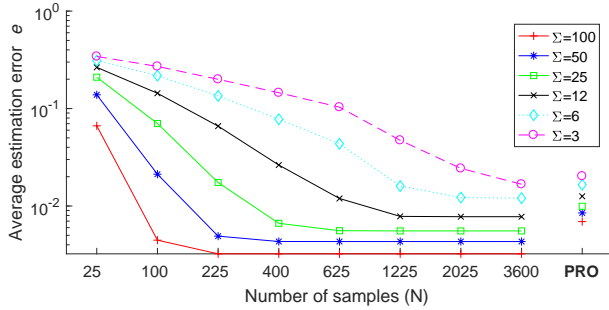
bilities of each cameras is within a range of 700m, i.e. cameras at opposite corners are not directly connected, and the data transmission is error free. Targets move in a  $500\text{m} \times 500\text{m}$  area and the state of each target  $j$  considers its 2D position and velocity:  $\mathbf{x}_j(k) = [x_j(k), y_j(k), \dot{x}_j(k), \dot{y}_j(k)]$ . The state transition matrix  $\Phi \in \mathcal{R}^{4 \times 4}$  is a first order motion model [12] with a process noise covariance matrix  $Q = \text{diag}([10 \ 10 \ 1 \ 1])$ . For observations representing 2D locations, we use a standard measurement matrix  $H_i^j \in \mathcal{R}^{2 \times 4}$  [14]; and noise covariance matrix  $R = \text{diag}([5 \ 5])$ .

#### 4.1. Detection quality

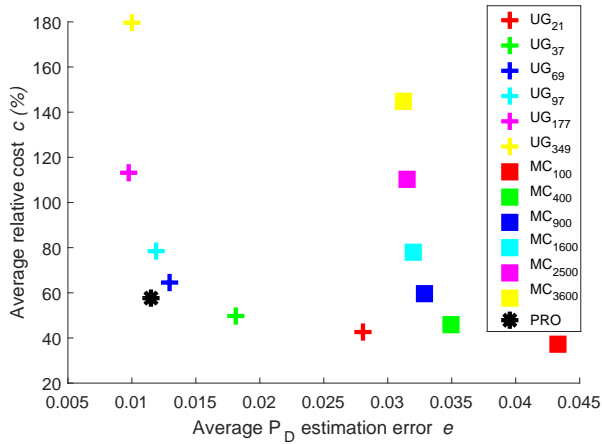
We estimate the detection quality  $P_D^{i,n}$  for  $M = 6 \cdot 10^5$  runs, i.e. for  $10^5$  random locations  $\mathbf{z}_i^n$  and six uncertainty hypotheses  $\Sigma = \{\Sigma_u = \text{diag}(100, 100)/2^u\}_{u=1}^6$  where  $\Sigma_u \in \mathcal{R}^{2 \times 2}$  is the measurement covariance employed by  $f(\mathbf{z}_i^n)$  to model the associated uncertainty of  $\mathbf{z}_i^n$  as  $\mathcal{N}(\mathbf{z}_i^n, \Sigma_u)$ . Note that  $\Sigma_u$  expands  $R$  to consider multiple observation accuracies of the detector.

We measure the average estimation error  $e = \frac{1}{M} |\bar{P}_D^{i,n} - \tilde{P}_D^{i,n}|$ , where  $\tilde{P}_D^{i,n}$  is the estimation and  $\bar{P}_D^{i,n}$  is the reference value found by evaluating all locations in  $\Omega_i$ . We also compare the relative computational cost  $c = 100 \frac{\text{cost}}{\text{cost}_R}$ , where  $\text{cost}$  and  $\text{cost}_R$  are, respectively, the total execution time (in MATLAB) of an estimation approach and the exhaustive integration, i.e. the reference estimation.

Fig. 4 shows the performance of the proposed approach and Quadrature-based Numerical Integration (QNI) [15], with a number of integration samples ranging from 25 to 3600. While the estimation error of QNI decreases when the number of samples increases (see Fig. 4a), its computational cost increases considerably (see Fig. 4b). Note that with  $N = 3600$  the cost is higher than



**Fig. 5.** Dependency of the estimation error on each uncertainty hypothesis  $\Sigma_u$  and number of samples.



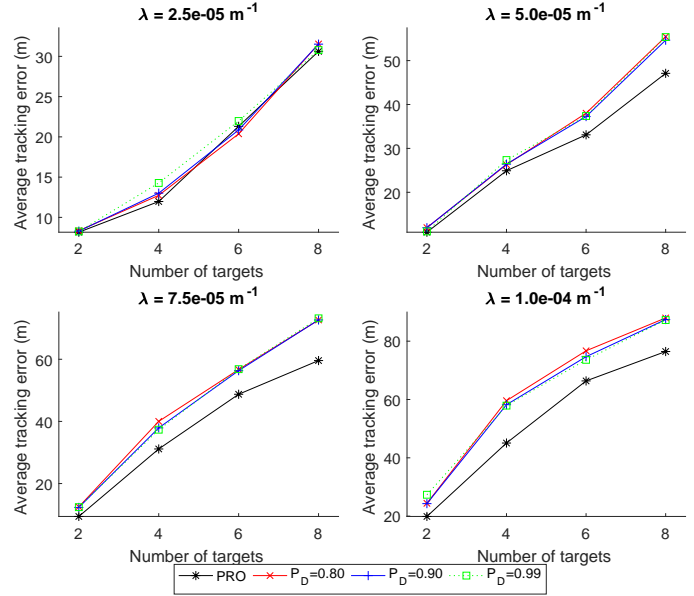
**Fig. 6.** Accuracy-complexity comparison for Importance Sampling (IS) approaches. KEY –  $UG_X$ : Uniform Grid IS with  $X$  samples.  $MC_X$ : Monte Carlo IS with  $X$  samples. PRO: Proposed method.

that of computing the reference value. Instead, the proposed method uses, on average, only 16 samples to estimate  $P_D$  and outperforms all QNIs. The accuracy of the proposed method is similar to that of QNI with 3600 samples and a cost between the cases with 625 and 1225 samples (reducing such cost  $\approx 2.5 \times$  to that of 3600 samples). Next, Fig. 5 shows the estimation accuracy when varying the uncertainty of the locations  $\Sigma_u$  of the detections.  $N$  has to increase to achieve accurate estimations for decreasing values in  $\Sigma_u$ .

Fig. 6 compares the proposed approach against two common importance sampling methods employing the uncertainty area (i.e. an ellipse) to get samples based on uniform grids (UG) [20] and Monte Carlo (MC) sample generation [21]. UG has low errors when considering a significant number of samples ( $> 69$ ) but the associated cost is higher than that of the proposed approach. The generation of samples by MC is computationally efficient whilst slightly decreasing the estimation error. The generation of samples by the proposed approach is highly efficient due to the refinement of importance sampling with quadrature-based rules. Overall, the proposed approach outperforms both UG and MC in terms of accuracy-cost trade-off.

#### 4.2. Tracking performance

We use the proposed approach to compute  $P_D$  for triangular FOVs for the MTIC distributed multi-target tracker [12] and compare it with the original version, which uses squared FOVs and a fixed  $P_D$



**Fig. 7.** Tracking error for the MTIC filter employing the proposed method and fixed values  $P_D \in \{.80, .90, .99\}$ .

value. Each camera is affected by false measurements (clutter) that we model with a Poisson distribution [14] defined by  $\lambda = r \cdot 10^{-5}$ , where the values of  $r$  are 2.5, 5.0, 7.5, and 10 (see as example Figure 3b). We also account for the common premise of prior detector knowledge [8, 9] by using three detection quality values  $P_D \in \{0.80, 0.90, 0.99\}$ . All the other MTIC parameters are set with the values defined by the authors [12].

As performance, we measure the tracking error between the final estimation and the ground-truth value of the target state. We set  $K = 40$  steps. We report the average error over 5600 runs (100 simulations for each clutter density,  $P_D$  value and number of targets  $N_t=2-8$ ). Fig. 7 compares the tracking performance of the proposed  $P_D$  computation and the fixed three  $P_D$  values. As the clutter level increases, false alarms reduce the successful association in PDA and increase the tracking error. The proposed approach considers real-valued  $P_D$  and observations  $\mathbf{z}_i^n$  with uncertain locations (i.e. close to the FOV limits) are better handled within the PDA used in MTIC.

## 5. CONCLUSIONS

We proposed to estimate target detection quality using quadrature-based integration and importance sampling. We determine the optimal number of samples for the integration by exploring the relationship between uncertainty models and the shapes of the FOVs of the camera. The proposed method uses importance sampling to guide the selection of samples and to increase the robustness to uncertain hypotheses. The proposed approach outperforms competitive methods in terms of accuracy and computational cost. Moreover, we demonstrated the use of the proposed approach to improve a state-of-the-art multi-target and multi-camera tracker under clutter using FOVs with triangular shapes.

As future work, we will explore the inclusion of appearance features within the proposal for  $P_D$  computation.

## 6. REFERENCES

- [1] H. Zhou, M. Taj, and A. Cavallaro, "Target detection and tracking with heterogeneous sensors," *IEEE Journal of Selected Topics in Signal Processing*, vol. 2, no. 4, pp. 503–513, Feb. 2008.
- [2] Y. Bar-Shalom, F. Daum, and J. Huang, "The probabilistic data association filter," *IEEE Control Systems Magazine*, vol. 29, no. 6, pp. 82–100, Dec. 2009.
- [3] S. Oh, S. Russell, and S. Sastry, "Markov chain monte carlo data association for multi-target tracking," *IEEE Trans. on Automatic Control*, vol. 54, no. 3, pp. 481–497, Mar 2009.
- [4] M. Breitenstein, F. Reichlin, B. Leibe, E. Koller-Meier, and L. Van Gool, "Robust tracking-by-detection using a detector confidence particle filter," in *Proc. of IEEE Int. Conference on Computer Vision (ICCV)*, Sept. 2009, pp. 1515–1522.
- [5] L. Kong, R. Dai, and Y. Zhang, "A new quality model for object detection using compressed videos," in *Proc. of IEEE Int. Conference on Image Processing (ICIP)*, Sept 2016, pp. 3797–3801.
- [6] P. Wang and Q. Ji, "Learning discriminant features for multi-view face and eye detection," in *Proc. of IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Jun. 2005, pp. 373–379.
- [7] J. Liebelt and C. Schmid, "Multi-view object class detection with a 3D geometric model," in *Proc. of IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Jun. 2010, pp. 1688–1695.
- [8] S. Rezatofghi, A. Milan, Z. Zhang, Q. Shi, A. Dick, and I. Reid, "Joint Probabilistic Data Association Revisited," in *Proc. of IEEE Int. Conference on Computer Vision (ICCV)*, Dec 2015, pp. 3047–3055.
- [9] D. Schulz, W. Burgard, D. Fox, and A. B. Cremers, "Tracking multiple moving targets with a mobile robot using particle filters and statistical data association," in *Proc. of IEEE Int. Conference on Robotics and Automation (ICRA)*, May 2001, vol. 2, pp. 1665–1670.
- [10] H. Aghajan and A. Cavallaro, *Multi-camera networks: principles and applications*, Academic press, 2009.
- [11] A. Mavrinnac and X. Chen, "Modeling coverage in camera networks: A survey," *International Journal of Computer Vision*, vol. 101, no. 1, pp. 205–226, Jan 2013.
- [12] A. Kamal, J. Bappy, J. Farrell, and A. Roy-Chowdhury, "Distributed multi-target tracking and data association in vision networks," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 38, no. 7, pp. 1397–1410, Jun 2016.
- [13] S. Katragadda, J. SanMiguel, and A. Cavallaro, "The costs of fusion in smart camera networks," in *Proc. of Int. Conference on Distributed Smart Cameras (ICDSC)*, Nov. 2014, pp. 1–6.
- [14] Y. Bar-Shalom, P. Willett, and X. Tian, *Tracking and data fusion*, YBS publishing, 2011.
- [15] D. Laurie, "Computation of gauss-type quadrature formulas," *Journal of Computational and Applied Mathematics*, vol. 127, no. 1, pp. 201–217, Jan. 2001.
- [16] A. Doucet, S. Godsill, and C. Andrieu, "On sequential monte carlo sampling methods for bayesian filtering," *Statistics and Computing*, vol. 10, no. 3, pp. 197–208, 2000.
- [17] C. Ding, B. Song, A. Morye, J. Farrell, and A. K. Roy-Chowdhury, "Collaborative sensing in a distributed PTZ camera network," *IEEE Trans. on Image Processing*, vol. 21, no. 7, pp. 3282–3295, 2012.
- [18] D. Dunavant, "High degree efficient symmetrical gaussian quadrature rules for the triangle," *Int. Journal for Numerical Methods in Engineering*, vol. 21, no. 6, pp. 1129–1148, 1985.
- [19] J. Lyness and R. Cools, "A survey of numerical cubature over triangles," in *Proc. of Symposia in Applied Mathematics*, Dec. 1994, vol. 48, pp. 127–150.
- [20] W. Press, *Numerical recipes: The art of scientific computing*, Cambridge university press, 3rd Edition, 2007.
- [21] C. Robert and G. Casella, *Monte Carlo statistical methods*, Springer Science & Business Media, 2013.