EUSIPCO 2013 1569746403

# PEDESTRIAN GROUP TRACKING USING THE GM-PHD FILTER

*Viktor Edman, Maria Andersson*

Div of Sensor Informatics
Dept of Sensor & EW Systems
Swedish Defence Research Agency
SE-581 11, Linköping, Sweden
viked065@student.liu.se
maria.h.andersson@liu.se

*Karl Granström, Fredrik Gustafsson*

Div of Automatic Control
Dept of Electrical Engineering
Linköping University
SE-581 83, Linköping, Sweden
karl@isy.liu.se
fredrik@isy.liu.se

## ABSTRACT

A GM-PHD filter is used for pedestrian tracking in a crowd surveillance application. The purpose is to keep track of the different groups over time as well as to represent the shape of the groups and the number of people within the groups. Input data to the GM-PHD filter are detections using a state of the art algorithm applied to video frames from the PETS 2012 benchmark data. In a first step, the detections in the frames are converted from image coordinates to world coordinates. This implies that groups can be defined in physical units in terms of distance in meters and speed differences in meters per second. The GM-PHD filter is a Bayesian framework that does not form tracks of individuals. Its output is well suited for clustering of individuals into groups. The results demonstrate that the GM-PHD filter has the capability of estimating the correct number of groups with an accurate representation of their sizes and shapes.

***Index Terms***— Multi target tracking, group target tracking, GM-PHD, groups.

## 1. INTRODUCTION

Multiple Target Tracking (MTT) in crowded scenes is a complex and difficult task. A crucial part for MTT is person detection and data association, where data association is the processes of recognizing the same person, among other persons, in consecutive frames. Typical techniques for single target state estimation include Kalman filtering, extended Kalman filtering and particle filtering, see e.g. [1–3]. Typical techniques for data association for multiple targets include the Joint Probabilistic Data Association Filter (JPDAF) and Multi Hypothesis Tracking (MHT), see e.g. [2]. In crowded scenes detections may not always be received from all persons in all frames because of occlusion. Therefore fewer tracks may be present than the actual number of persons. Moreover, the tracks may easily switch identities. In crowded scenes group tracking is often a better and more effective alternative since the handling of the different objects (which are one or more groups) can be made easier in the tracking algorithm and, moreover, we do not always need to track and identify each person in the groups. Group tracking has been investigated in several studies and for several applications, see e.g. [4–10].

A related problem to group tracking is the tracking of so-called extended targets. An extended target is a target that potentially gives rise to more than one measurement per time step. Solutions for multiple extended target tracking, e.g. [11–14], can be used for multiple group tracking.

Approaches for solving group tracking can roughly be divided into the following [2]:

1. Group tracking without individual tracks;
2. Group tracking with simplified individual tracks;
3. Individual target tracking which is supplemented by group tracking.

The most suitable approach largely depends on the application. In crowded scenes, with many potentially false detections and clutter, 1. or 2. would probably be the most practical approaches since tracks of all individuals within the group will be difficult to initiate and maintain.

Group tracking uses the same processes as conventional tracking methods, i.e. detection, association and prediction. An additional step required for group tracking is the representation of the group, in the form of shape and size. The shape and size of the group can also be used to estimate the behavior of the group. This is done in for example [15], using clustering techniques, and in [16], using the PHD filter. The behavior of the group is in these studies represented by group activity (e.g. fights), merge and split.

In this paper we continue the work from [15] and investigate the advantage of the GM-PHD filter for handling, in video surveillance, a varying number of groups over time. The novelty of this paper is that we investigate the GM-PHD filter together with the detection step (including the conversion from image coordinates to world coordinates) and that

we test the approach on real video data. In this way we can produce real detections from the video data to use as input data to the GM-PDH filter, and thereby consider also the important detection uncertainties.

## 2. METHOD AND APPROACH

The proposed approach is outlined below.

1. For each image frame, pedestrians are detected by a state of the art method [17, 18]. Each pedestrian is represented with a rectangle, and we pick the mid point of the lower side as an estimate of each pedestrian's footprint. The output of the algorithm is a point $p_I$ in image coordinates. Future work will study if a realistic covariance matrix $P_I$ can be derived as well.

2. These points are transformed to world coordinates $p_W$ with covariance $P_W$. This assumes that the video camera is placed on an elevated position, and that a terrain elevation map is available for the scene.

3. The GM-PHD filter represents the multi target information with a Gaussian Mixture (GM) approximation of the PHD intensity $v_k(x)$ over the state space $x_k$,

$$v_k(x) = \sum_i w_k^{(i)} \mathcal{N}\left(x_k; \mathbf{m}_k^{(i)}, P_k^{(i)}\right). \qquad (1)$$

   In this paper we have $x_k = [\mathsf{p}_k^T, \mathsf{v}_k^T]^T$, where $\mathsf{p}_k$ is the position and $\mathsf{v}_k$ is the velocity at time $k$. It is important to note that the modes in the PHD intensity $v_k(x)$ do not correspond to individuals, unlike classical filterbank target tracking methods, see e.g. [1]. Instead the PHD intensity is defined by the property that the integral

$$\int_{A,V} v_k\big([\mathsf{p}^T, \mathsf{v}^T]^T\big) d\mathsf{p}\, d\mathsf{v} \qquad (2)$$

   is the expected number of pedestrians within the area $A$ and velocity interval $V$. The GM-PHD filter approximates the Bayesian solution of this using the detections $p_W$ in world coordinates as the only input.

   Though we here study a single camera application, the PHD filter framework can also handle multiple sensors. Theory and application on PHD filters for multiple sensors are discussed and presented in for example [19–21].

4. The final step is to apply a clustering algorithm to the GM-PHD filter output. The GM representation is particularly well suited for clustering. The main idea in the clustering for this application is to find level curves separating the groups in both position and velocity. The integral of the GM-PHD density within each contour estimates the size of the group.

It should be stressed that all steps are within a sound Bayesian framework, where the approximation in the algorithm can be arbitrarily small by increasing the size of the GM. It is only the final clustering step that is ad-hoc, but it has no memory. The main challenge is the tuning part where the design parameters in the filter are chosen.

## 3. DETECTION OF PEDSTRIANS

### 3.1. Detection in Image Frames

For detection of pedestrians in the dataset the methods and code presented by Piotr Dollár [17, 18] are used. The detection algorithm uses integral channel features for extracting pedestrians from a single image, no prior information is needed for the detections. Dollár concludes that this method outperforms for instance the method based on histogram of oriented gradients. The detection algorithm was run with the settings: `resize=1.2` ,`fast=0`, `modelNm=ChnFtrs01`.

Partly due to the lack of prior knowledge the algorithm has difficulties detecting pedestrians that are partly or fully obscured by other objects. This gives rise to missed detections. This is handled by the PHD filter by the parameter $p_D$, which is assumed to be known for each scenario.

The algorithm returns a bounding box for each detected pedestrian. It is not in the scope of this paper improving this algorithm. It is only used for extracting measurement from the dataset.

### 3.2. Camera Calibration

The data from PETS 2012 [22] includes a camera calibration file. The file contains different calibration parameters that have been determined by using *Tsai camera calibration model* [23]. These parameters can be used to transform image coordinates $(x_f, y_f)$ to ground plane coordinates $(\mathsf{x}, \mathsf{y}, \mathsf{z})$.

The first step is to transform the image coordinates $(x_f, y_f)$ into distorted image coordinates $(x_d, y_d)$.

$$x_d = d_x(x_f - C_x)/s_x, \qquad y_d = d_y(y_f - C_y), \qquad (3)$$

where $d_x, d_y$ are center to center distance between adjacent sensor elements in $x$ and $y$ direction respectively, $C_y, C_x$ are coordinates of center of radial lens distortion and $s_x$ is a scale factor compensating for uncertainty imperfections in hardware timing for scanning and digitisation.

The second step is to transform the distorted coordinates into undistorted image coordinates $(x_u, y_u)$.

$$x_u = x_d(1 + \kappa r^2), \qquad y_u = y_d(1 + \kappa r^2). \qquad (4)$$

where $r = \sqrt{x_d^2 + y_d^2}$ and $\kappa$ is the radial lens distortion coefficient.

### 3.3. Conversion to Ground Plane

Tracking objects in the image plane is possible, but the drawback is that physical motion of pedestrians is harder to model in the image plane. Further, clustering is easier to perform in physical quantities. Instead of tracking in the image plane the goal is to follow both individuals and groups in the ground plane, i.e. in world coordinates. Hence the center point for the lower edge of each bounding box is transformed into world coordinates, which are used as measurements. This is easily done by assuming that all targets move in the ground plane defined by $\mathsf{z}(\mathsf{x},\mathsf{y})$ given by a terrain elevation map (TEM).

The transformation in general is given by the following system of equations:

$$\begin{bmatrix} x_u z_c/f & y_u z_c/f & z_c \end{bmatrix}^T = R \begin{bmatrix} \mathsf{x} & \mathsf{y} & \mathsf{z}(\mathsf{x},\mathsf{y}) \end{bmatrix}^T + T, \quad (5)$$

where $f$ is the focal length, $z_c$ is the camera's $z$-coordinate which is unknown, $R$ is a rotation matrix, and $T = [T_x, T_y, T_z]^T$ is a translation vector. The solution for a flat world $\mathsf{z}(\mathsf{x},\mathsf{y}) = 0$ is given by

$$\mathsf{x} = \frac{(T_x - x_u T_z/f)(y_u R_{3,2}/f - R_{2,2}) - (x_u R_{3,2}/f - R_{1,2})(T_y - y_u T_z/f)}{(x_u R_{3,1}/f - R_{1,1})(y_u R_{3,2}/f - R_{2,2}) - (x_u R_{3,2}/f - R_{1,2})(y_u R_{3,1}/f - R_{2,1})} \quad (6)$$

$$\mathsf{y} = \frac{(x_u R_{3,1}/f - R_{1,1})(T_y - y_u T_z/f) - (T_x - x_u T_z/f)(y_u R_{3,1}/f - R_{2,1})}{(x_u R_{3,1}/f - R_{1,1})(y_u R_{3,2}/f - R_{2,2}) - (x_u R_{3,2}/f - R_{1,2})(y_u R_{3,1}/f - R_{2,1})}. \quad (7)$$

This solution provides a good initial value for non-flat worlds, where a few gradient or Gauss-Newton steps should suffice to improve the solution.

## 4. GAUSSIAN MIXTURE PROBABILITY HYPOTHESIS DENSITY FILTER

The PHD filter is a rigorous Bayesian solution to the multi-target tracking problem [24, 25]. Its Gaussian Mixture implementation, called the GM-PHD filter, is presented in [26]. Below we give the modelling choice that were made in this work. Refer to [26] for the PHD-filter equations and pseudo code. The state vector $\mathbf{x}$ contains four states: position in both $x$- and $y$-direction, and corresponding velocities. The sampling time is $T_s = 1/7$.

### 4.1. Initialization

The GM-PHD intensity is initialized with $J_0 = 4$ components

$$v_0(\mathbf{x}) = \sum_{i=1}^{J_0} w_0^{(i)} \mathcal{N}\left(\mathbf{x}; \mathbf{m}_0^{(i)}, P_0^{(i)}\right), \quad (8a)$$

$$w_0^{(1)} = w_0^{(2)} = w_0^{(3)} = w_0^{(4)} = 1, \quad (8b)$$

$$m_0^{(1)} = \begin{bmatrix} -11.2197 & -13.1848 & 0 & 0 \end{bmatrix}^T, \quad (8c)$$

$$m_0^{(2)} = \begin{bmatrix} -11.1650 & -14.1883 & 0 & 0 \end{bmatrix}^T, \quad (8d)$$

$$m_0^{(3)} = \begin{bmatrix} -9.2323 & -13.8840 & 0 & 0 \end{bmatrix}^T, \quad (8e)$$

$$m_0^{(4)} = \begin{bmatrix} -7.9414 & 4.3781 & 0 & 0 \end{bmatrix}^T, \quad (8f)$$

$$P_0^{(1)} = P_0^{(2)} = P_0^{(3)} = P_0^{(4)} = \mathrm{diag}\,(0.1, 0.1, 1, 1). \quad (8g)$$

### 4.2. Prediction

For surviving targets the probability of survival is set to $p_S = 0.99$. The motion of the targets is modelled according to a constant velocity model. The uncertainty of the model is modelled as white Gaussian noise with covariance matrix $Q_k = G_k \mathrm{diag}\,(1,1)\, G_k^T$ where

$$G_k = \begin{bmatrix} \frac{T_s^2}{2} & 0 & T_s & 0 \\ 0 & \frac{T_s^2}{2} & 0 & T_s \end{bmatrix}^T. \quad (9)$$

The spontaneous birth PHD has $J_\gamma = 3$ components

$$\gamma_k(\mathbf{x}) = \sum_{i=1}^{J_\gamma} w_\gamma^{(i)} \mathcal{N}\left(\mathbf{x}; \mathbf{m}_\gamma^{(i)}, P_\gamma^{(i)}\right) \quad (10a)$$

$$w_\gamma^{(1)} = 0.01,\ w_\gamma^{(2)} = 0.001,\ w_\gamma^{(3)} = 0.0001, \quad (10b)$$

$$m_\gamma^{(1)} = \begin{bmatrix} -8 & 6 & 0 & 0 \end{bmatrix}^T \quad (10c)$$

$$m_\gamma^{(2)} = \begin{bmatrix} -10 & -15 & 0 & 0 \end{bmatrix}^T \quad (10d)$$

$$m_\gamma^{(3)} = \begin{bmatrix} 15 & -8 & 0 & 0 \end{bmatrix}^T \quad (10e)$$

$$P_\gamma^{(1)} = P_\gamma^{(2)} = P_\gamma^{(3)} = \mathrm{diag}\,(1,1,1,1). \quad (10f)$$

This means that new targets are modelled as being likely to appear at the places where the road intersects the camera's field of view. Target spawning is omitted in this work.

### 4.3. Measurement Update

The target detections are modelled as linear measurements of the target position. The uncertainty of the measurements is modelled as Gaussian white noise with covariance $R = \mathrm{diag}\,(0.5, 0.5, 0.5, 0.5)$. The probability of detection is set to $p_D = 0.7$ and the parameter modelling the clutter is set to $\kappa = 10^{-8}$.

### 4.4. Merging and Pruning

Pruning and merging is employed to keep the number of PHD components at a tractable level. After the measurement update, components with weight $w_k^{(i)} < 10^{-5}$ are pruned. Next components with Mahalanobis distance less than $U = 2$ from eachother are merged. If there still are too many components after merging only the $J_{max} = 100$ components with the highest weights are saved.

## 5. CLUSTERING OF GROUPS

After pruning and merging in the GM-PHD filter, the Gaussian components are divided into groups. The division is done by calculating the euclidean distance and difference in velocity for all combinations of Gaussian components above a

given weight. If the distance and difference in velocity between two components are below some thresholds $T_p = 2[m]$ and $T_v = 1[m/s]$, they are considered to be connected. All components that in some way are connected are considered to be a part of the same group.

For each group $i$ the GM-PHD surface is calculated as

$$v_{k|k}^{(i)}(\mathbf{x}) = \sum_{j=1}^{J_{i,k|k}} w_{k|k}^{(i,j)} \mathcal{N}\left(\mathbf{x}; \mathbf{m}_{k|k}^{(i,j)}, P_{k|k}^{(i,j)}\right), \tag{11}$$

and intersected at a desired height, which in this study is 0.1. This intersection is interpreted as an approximation of the groups' shapes and sizes. To estimate the number of members in a group the corresponding weights are summed up according to

$$\hat{N}_{k|k}^{(i)} = \sum_{j=1}^{J_{i,k}} w_{k|k}^{(i,j)}. \tag{12}$$

## 6. EXPERIMENTS

This section presents results from the experiments that have been performed. The dataset used for the group tracking is *Flow Analysis and Event Recognition*, marked 13:57 using camera view 1, from the PETS 2012 dataset [22]. In the scenario several groups of people move along a road from one edge of the image to the other. All groups move in the same direction (right to left in the image), with the exception of a single person which is moving in the opposite direction.

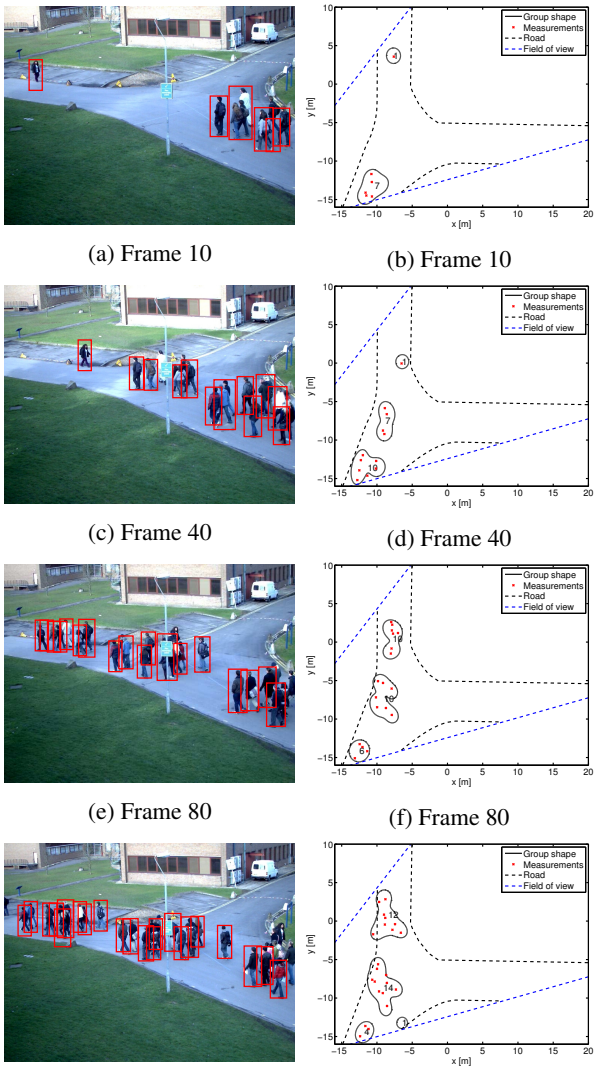
Figure 1 displays the estimated shape of the groups and the estimated number of individuals in respective group. The estimated number of individuals in the whole scene and an estimated number of groups can be seen in Figure 2. The results can also be seen in a video at `youtu.be/aAz3poW49CU`.

## 7. CONCLUSIONS

The GM-PHD filter provides a computational engine suitable for post-processing of the information in image detections. We applied a simple clustering algorithm to its output, which can readily solve the group clustering in physical units. That is, we can define a group as individuals closer to each other than two meters and with a velocity within one meter per second.

Further, the GM-PHD surface presents a nice visualisation of the groups' estimated extensions, suitable as a high-level presentation for manual operators in surveillance applications. It is also possible to predict groups that are approaching, and thus give an early warning and potential conflict alerts.

The low probability of detection implied by image detection algorithms is a slight problem for the GM-PHD filter. If a group is obscured by another group for several frames the group will disappear from the filter. Consequently, the estimated number of individuals is a rough approximate which can be seen in Figure 2. However, this estimate is significantly better than taking the number of detections as an estimate of the number of individuals. Future work will develop refined merging and pruning steps for the GM-PHD filter. Another possible remedy is provided by the cardinalized GM-PHD filter.



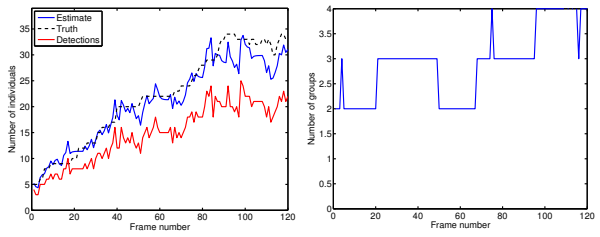

(a) Frame 10 (b) Frame 10

(c) Frame 40 (d) Frame 40

(e) Frame 80 (f) Frame 80

(g) Frame 120 (h) Frame 120

**Fig. 1**: (a,c,e,g) The scene with rectangles denoting detections in frame number 10, 40, 80 and 120. (b,d,f,h) Estimated groups in frame number 10, 40, 80 and 120. The numbers adjacent to the groups are the estimates of the numbers of individuals in the respective groups, the measurements are denoted with crosses, and the dashed lines are the camera's field of view, and the edges of the road, respectively.




(a) (b)

**Fig. 2**: Results from experiment. (a) Plot displaying total number of estimated individuals in the scene compared to the actual number of individuals and the number of detections. (b) Plot displaying the estimated number of groups in image over frames.

## 8. REFERENCES


[1] Y. Bar-Shalom, P. K. Willett, and X. Tian, *Tracking and data fusion, a handbook of algorithms*, YBS, 2011.

[2] S. S. Blackman and R. Popoli, *Design and analysis of modern tracking systems*, Artech radar library. Artech House, Norwood, MA, USA, 1999.

[3] F. Gustafsson, F. Gunnarsson, N. Bergman, U. Forssell, J. Jansson, R. Karlsson, and P.J. Nordlund, "Particle filters for positioning, navigation, and tracking," *IEEE Transactions on Signal Processing*, vol. 50, no. 2, pp. 425–437, 2002.

[4] W. Konle, "Tracking of aircraft groups in an operational air surveillance system.," *2011 Proceedings of the 14th International Conference on Information Fusion (FUSION)*, pp. 1–6, 2011.

[5] A. Swain and D. Clark, "The single-group PHD filter: An analytic solution.," in *Fusion 2011 - 14th International Conference on Information Fusion*, 2011.

[6] M. Baum, B. Noack, and U.D. Hanebeck, "Extended object and group tracking with elliptic random hypersurface models.," in *13th Conference on Information Fusion, Fusion 2010*, 2010.

[7] B. Zhan, D. N. Monekosso, P. Remagnino, S. A. Velastin, and L.-Q. Xu, "Crowd analysis: a survey," *Machine Vision and Applications*, vol. 19, no. 5, pp. 345–357, 2008.

[8] D. Clark and S. Godsill, "Group target tracking with the gaussian mixture probability hypothesis density filter," in *Intelligent Sensors, Sensor Networks and Information, 2007. ISSNIP 2007. 3rd International Conference on*. IEEE, 2007, pp. 149–154.

[9] J. Rosswog and K. Ghose, "Detecting and tracking coordinated groups in dense, systematically moving, crowds," *SDM12*, pp. 1–11, 2012.

[10] S. J McKenna, S. Jabri, Z. Duric, A. Rosenfeld, and H. Wechsler, "Tracking groups of people," *Computer Vision and Image Understanding*, vol. 80, no. 1, pp. 42–56, 2000.

[11] K. Granström, C. Lundquist, and U. Orguner, "A Gaussian mixture PHD filter for extended target tracking," in *Proceedings of International Conference on Information Fusion (FUSION)*, Edinburgh, UK, July 2010.

[12] K. Granström, C. Lundquist, and U. Orguner, "Extended Target Tracking using a Gaussian Mixture PHD filter," *IEEE Transactions on Aerospace and Electrical Systems*, vol. 48, no. 4, pp. 3268–3286, Oct. 2012.

[13] K. Granström and U. Orguner, "A PHD filter for tracking multiple extended targets using random matrices," *IEEE Transactions on Signal Processing*, vol. 60, no. 11, pp. 5657–5671, Nov. 2012.

[14] C. Lundquist, K. Granström, and U. Orguner, "An extended target CPHD filter and a gamma Gaussian inverse Wishart implementation," *IEEE Journal of Selected Topics in Signal Processing, Special Issue on Multi-target Tracking*, vol. 7, no. 3, pp. 472–483, June 2013.

[15] M. Andersson, F. Gustafsson, L. St-Laurent, and D. Prevost, "Recognition of anomalous motion patterns in urban surveillance," *IEEE Journal of Selected Topics in Signal Processing*, vol. 7, no. 1, pp. 102–110, Feb. 2013.

[16] A. Carmi, F. Septier, and S. J. Godsill, "The gaussian mixture MCMC particle algorithm for dynamic cluster tracking," *Automatica*, 2012.

[17] P. Dollár, Z. Tu, P. Perona, and S. Belongie, "Integral channel features," in *BMVC*, 2009.

[18] P. Dollár, S. Belongie, and P. Perona, "The fastest pedestrian detector in the west," in *BMVC*, 2010.

[19] R Mahler, "The multisensor PHD filter 1: General solution via multitarget calculus," in *SPIE*, 2009.

[20] R Mahler, "The multisensor PHD filter 11: Erroneous solution via 'poisson magic'," in *SPIE*, 2009.

[21] E. Delande, P. Duflos, P. Vanheeghe, and D. Heurguier, "Multisensor PHD construction and implementation by space partitioning," in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2011.

[22] University of Reading, "PETS 2012 dataset S1: Person count and density estimation," Jan. 2012.

[23] R. Tsai, "A versatile camera calibration technique for high-accuracy 3d machine vision metrology using off-the-shelf TV cameras and lenses," *Robotics and Automation, IEEE Journal of*, vol. 3, no. 4, pp. 323 –344, Aug. 1987.

[24] R.P.S. Mahler, "Multitarget Bayes filtering via first-order multi target moments," *IEEE Transactions on Aerospace and Electrical Systems*, vol. 39, no. 4, pp. 1152–1178, Oct. 2003.

[25] R.P.S. Mahler, *Statistical Multisource-Multitarget Information Fusion*, Artech House, Norwood, MA, USA, 2007.

[26] B.-N. Vo and W.-K. Ma, "The Gaussian mixture probability hypothesis density filter," *Signal Processing, IEEE Transactions on*, vol. 54, no. 11, pp. 4091 –4104, Nov. 2006.