Recent Advances in Extended and Group Objects Tracking

Lyudmila Mihaylova and Waqas Aftab
Automatic Control and Systems Engineering, University of Sheffield
Amy Johnson Building, Mappin Street, Sheffield S1 3JD
United Kingdom

Email: l.s.mihaylova@sheffield.ac.uk, waftab1@sheffield.ac.uk

ABSTRACT

We live in an era of increasing data and information from multiple sensors. The multiple complementary types of sensors introduce a variety of challenges, especially in systems with different level of autonomy, such as Unmanned Aerial Vehicles and surveillance systems. Autonomous systems require quick situation awareness, including tracking of the location and size of the objects of interest, for instance extended and groups which give rise to multiple measurements. Examples of extended objects are pedestrians, convoys of vehicles and clouds of bio-chemical contaminants. Most of the current approaches rely on well-defined mathematical models. However, the changes both in groups and extended objects dynamics and in the environment require flexible approaches able to learn and adapt to the changes. Hence, this work overviews the state-of-the-art approaches and focusses on data-driven approaches such as Gaussian processes for spatio-temporal representations of extended objects and groups. We share our vision for future trends in this area .

1.0 INTRODUCTION

Extended object tracking (EOT) encompasses estimation of the kinematic states and the extent of an object of interest using sensor measurements. Group object tracking (GOT) deals with the estimation of the kinematic states and the group shape which comprises multiple entities [1]. Groups are structured objects, formations of entities moving in a coordinated manner, following a pattern. The number of group entities varies over time since entities can enter a scene, or disappear at random times. Both groups and extended object give raise to multiple measurements and hence lead to data origin measurement uncertainty. The sensor measurements are utilised to estimate the shape, size, volume, orientation and other parameters that gives valuable information regarding the spatial objects of interest. This additional information is useful for various civilian and military systems as it helps in object classification, intent inference and prediction. In applications with dense point objects, e.g. a formation of objects, being reported by multiple noisy and biased sensors, tracking of individual objects becomes computationally expensive. Moreover, due to high uncertainty in the individual object motion and the measurement origin (objects are in close vicinity), the performance of the data association algorithms degrade to a level that it becomes difficult to maintain trajectories. Hence, GOT methods provide solutions to both of the above-mentioned problems, based on the fact that when objects are close to each other and have similar kinematics, then the average group kinematics, the shape and in some cases the number of objects provides a better situational awareness than individual object tracking. Additionally, it also reduces the computational expense of the data association algorithms.

As is the case with multiple target tracking (MTT), EOT and GOT methods have been applied to diverse fields such as medical, biological, chemical, urban traffic management (both indoors and outdoors), in security systems, autonomous vehicles (land, sea, air), robotics, to name a few. The algorithms primarily serve the autonomous part of these systems. Although, GOT concepts were introduced long time ago [2], they have been applied extensively mainly recently to real systems. There is an exponential growth in the systems employing EOT and GOT during the last two decades.

Similar problems have been studied by the computer vision community and are called visual object tracking (VOT). The challenges are different although the methods overlap in some areas. In VOT, the measurement data is typically images obtained using a camera (red, green, blue (RGB), grayscale, infrared (IR), multi-spectral and hyperspectral cameras). As a result, in most cases, a very precise contour of the object is available at all times. Comparatively, in EOT and GOT applications, only part of the object contour or surface is available at any particular scan.

The EOT and GOT research has been surveyed extensively in two overview papers [1] and [3]. This paper provides a survey of recent methods. The structure of the paper is as follows: Section 2 discusses the challenges that EOT and GOT face and presents a taxonomy of the main methods developed to cope with them. Section 3 describes the main concepts of the recently developed machine learning methods – Gaussian processes for EOT and GOT. Section 4 discusses typical datasets and performance measures used to evaluate the performance of methods for tracking of spatial objects. Section 5 concludes the paper and discussed future work.

2.0 CHALLENGES AND METHODS

The EOT and GOT deal with estimation of the object kinematics and extent from the sensor data. The estimation problem is complex as the sensor data is noisy, the object motion is unknown and the measurement origin is uncertain (it can be clutter). Additionally, in each time sample a portion of the object is detected by the sensor (instead of the complete object), the association of the measurement to the reflection point on the object is unknown and the reflection points change from sample to sample. The situation worsens in the presence of multiple closely situated objects. The object's shape is irregular or non-rigid and changes in time.

Typically, in the case of single EOT, the objective is subdivided into estimation of the kinematics of the centre of the object (COO) and the shape is estimated with respect to the COO location. The estimates are correlated as the error in the COO estimation propagates in the shape estimate. The COO kinematics estimation is similar to point object tracking, hence the focus of the EOT research has been the extent estimation. The extent estimation is done using two independent models, namely the shape model and the shape kinematics model. Various shape models such as based on Poisson point process distributions [4], rectangular [5], stick [6], random matrix (ellipse) [7] models have been proposed which estimate the shape as a basic geometric shape. Recently, complex shape models have been proposed such as random hypersurface (RHM) [8], a Gaussian process (GP) approach [9], a mixture of ellipsoids, B-splines [10] and others for irregular shapes. The GP method, being non-parametric, is flexible and provides better performance as compared with the RHM [9]. The track before detect extension of EOT has been proposed in [11] and improved upon in [12].

In the case of multiple extended objects, additional challenges to data association arise, also when the extended objects split or merge. Then two-step approaches have been proposed based on measurement clustering or partitioning followed by data association. The performance of measurement clustering approach is satisfactory for tracking distant objects. In the cases of closely situated objects or objects with crossing trajectories, the measurement clustering is not enough to resolve measurement ambiguities and then measurements partitioning is done. The measurement set partitioning deals with determining all the subsets of the measurement within a random finite set (RFS) [13]. A data association algorithm then provides the best measurement set to the object association based on the set to object state RFS likelihood. A lot of research has being done to determine the most probable partitions in real-time in order to reduce the computational complexity of the subsequent data association step. Some recent two-step based methods for EOT and GOT are as follows: the box particle filter (BPF) [14] and the Gaussian process convolution particle filter (GPCPF) [15], the multiple hypothesis tracker (MHT) based methods such as probabilistic MHT [16], the probabilistic data association (PDA) based methods such as joint PDA [17], Markov chain Monte Carlo methods [18] and the random finite sets (RFS) based methods such as probability hypothesis density (PHD) filters [19], multi-Bernuolli (MB) filters [20]. A one-step data association approach is proposed in [21]. The merging and splitting of extended and group objects has been studied but is not fully resolved.

Some of the main types of methods developed for EOT and GOT problems are summarised in Figure 1. However, the presented taxonomy is not exhaustive in terms of methods or applications. Although machine learning methods have been used for processing the data from soft sensors, these methods have also been gaining popularity in the hard sensor processing, recently. The hard sensors can be classified into conventional (active e.g. radar, LiDAR, IR and passive e.g. optical, etc.) and non-conventional (chemical, biological) types. The detection and tracking of objects using video data is done by pre-processing the measurements (detection) followed by the visual object tracking.

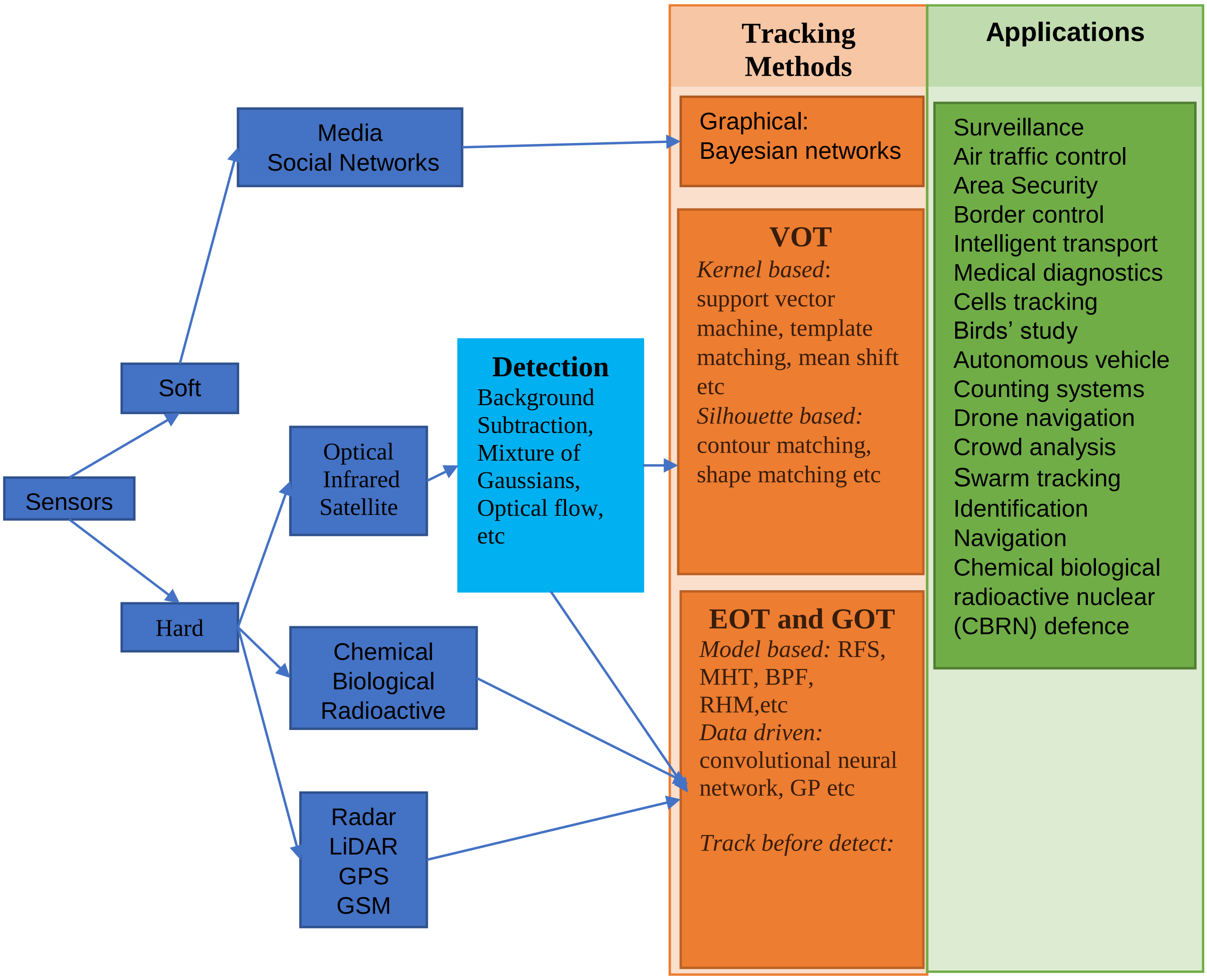

Figure 1. Taxonomy of methods for EOT and GOT and their applications.

This Figure shows the EOT and GOT methods and their applications based on hard and soft sensors.

3.0 DATA DRIVEN METHODS

Data driven methods relates to methods that make predictions of phenomena and events based on data analysis and interpretations rather than on models. These are also called Machine learning methods and have a big potential in dealing with high dimensional systems and large data. In the context of EOT and GOT, the GP and convolutional neural networks (CNN) are promising methods. Other approaches have also been developed for applications such as for pedestrian detection in autonomous vehicles in [22].

3.1 EOT and GOT Using Gaussian Process Approaches

Machine Learning (ML) encompasses data driven approaches that have been applied to diverse systems for various purposes: classification, inference, tracking, anomaly detection, to name a few. GP methods [23] are emerging ML methods but their application to EOT and GOT is still limited. In [9] a GP framework for EOT and GOT is proposed where the object shape is modelled using a GP. This is one of the few approaches capable of estimating a complex shape, compared with most of the methods that estimate the object as a basic geometrical shape that is stick, circle, rectangle or ellipse. GP is a powerful framework which has also been used for other complex object tracking problems from data association of multiple point objects to overlapping mixture of Gaussian process (OMGP) [24] for data association. Most of the ML methods are computationally expensive and so is the Gaussian process inference. This means that, for problems where the input space is unbounded, e.g. for time series, the computational expense of the Gaussian process based method will increase and at some point the current computational resources will not be enough to provide the solution in real-time. In [25], a recursive filtering and smoothing solution to the Gaussian process inference has been proposed, which can be used to determine real-time Gaussian process inference.

3.1.1 Background

A Gaussian process (GP) is a stochastic process used to map a nonlinear function from an input space to an output space. It is different from model-based techniques, where a distribution over parameters is used for estimation and decision making purposes. A GP models a distribution over functions in a nonparametric way and thus provides more flexibility. A GP is defined by a mean and a covariance kernel. The mean models the mean of the mapping function and the covariance kernel represents the correlation among the GP inputs. A quick introduction to the GP and on the choice of mean and the covariance kernel is provided in [23]. The hyperparameters, the parameters of the mean and covariance kernel, can be optimised by maximising the likelihood of a given input-output data and this process is called learning. The given data, also called the training data, and the learned hyperparameters can be used to predict the mean and the prediction uncertainty using a GP regression.

Suppose a GP models the nonlinear function f from a random input a to a random output b. Here we are considering mapping from a scalar (one dimensional, 1D) input to a scalar (1D) output for simplicity. The GP regression, given below, is applicable to multiple dimensional inputs and multiple dimensional outputs as well

$$b = f(a), \quad f(a) \sim GP\big(\mu(a), k(a, a')\big), \qquad (1)$$

where $\mu(\boldsymbol{a})$ represents the mean and $\boldsymbol{k}(\boldsymbol{a}, \boldsymbol{a}')$ represents the covariance kernel of the GP. The covariance kernel takes two input samples (these can be same or different, $\boldsymbol{a}'$ represents the order of the input) and depending on the covariance kernel function gives an output value. Suppose, the output $\boldsymbol{b}$ is observed at $\boldsymbol{n}$ different input locations, which can be modelled using the following equation:

$$\boldsymbol{y} = \mathbf{f}(\mathbf{a}) + \mathbf{v}, \qquad (2)$$

where $\boldsymbol{y} = [y_1, y_2, \dots, y_n]^T$ represents the measurement vector corresponding to input vector $\boldsymbol{a} = [a_1, a_2, \dots, a_n]^T$, $\boldsymbol{f} = [f(a_1), f(a_2), \dots, f(a_n)]^T$ represents the corresponding function vector, $[.]^T$ represents a vector transpose, $\boldsymbol{v} \sim \mathbb{N}(0, \sigma^2 \boldsymbol{I}_n)$ represents the independent identically distributed (i.i.d.) Gaussian measurement noise vector with variance σ^2 and $\boldsymbol{I}_n$ represents an n-dimensional identity matrix. Given n input-output data pairs, the GP regression can be used to determine the output at unknown input location vector $\boldsymbol{a}^*$ as given below:

$$\mu(\boldsymbol{a}^*) = \boldsymbol{C}_{a^*a}(\boldsymbol{C}_{aa} + \sigma^2 \boldsymbol{I}_n)^{-1} \boldsymbol{f}(\boldsymbol{a}), \quad (3)$$

$$\boldsymbol{C}(\boldsymbol{a}^*) = \boldsymbol{C}_{a^*a^*} - \boldsymbol{C}_{a^*a} (\boldsymbol{C}_{aa} + \sigma^2 \boldsymbol{I}_n)^{-1} \boldsymbol{C}_{aa^*}, \quad (4)$$

where $\mu(\boldsymbol{a}^*)$ represents the predicted mean output vector, $\boldsymbol{C}(\boldsymbol{a}^*)$ represents the corresponding output covariance matrix, $(.)^{-1}$ denotes a matrix inverse operation and $\boldsymbol{C}_{pq}$ represents the GP covariance matrix between the two input vectors $\boldsymbol{p}$ and $\boldsymbol{q}$ and is calculated as given below:

$$\boldsymbol{C}_{pq} = \begin{pmatrix} k(p, q_1) & k(p_1, q_2) & \cdots & k(p_1, q_l) \\ \vdots & & \ddots & \vdots \\ k(p_m, q_1) & k(p_m, q_2) & \cdots & k(p_m, q_l) \end{pmatrix}, \quad (5)$$

where the length of $\boldsymbol{p}$ and $\boldsymbol{q}$ is l and m, respectively.

3.1.2 Gaussian Process Convolution Particle Filter for Multiple EOT and GOT

The Gaussian Process Convolutional Particle filter (GPCPF) [15] tracks multiple irregularly shaped objects moving through clutter using surface measurements. The object shape or contour tracking complexity is increased due to the surface measurements, as the measurements do not necessarily represent the object contour. The GPCPF provides flexibility, unlike other approaches, and does not require the prior knowledge of the statistical properties of the object measurements. The tracking problem is formulated as a state space model. The state vector consists of the states belonging to the COO and the extent. The extent states are modelled using a Gaussian process [15] as shown in Fig. 2.

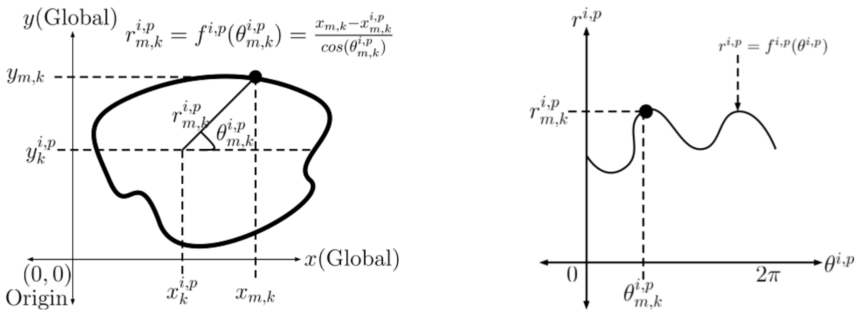

Figure 2: Gaussian process (GP) model for the object extent.

This figure shows a GP model for the extent proposed in [9]. The left axis show the object (thick solid line) and a measurement (dot). The measurement is received in global frame at coordinates $(x_{m,k}, y_{m,k})$. The measurement coordinates in the object frame with origin at the COO that is $(x^{i,p}_k, y^{i,p}_k)$ are $(r^{i,p}_{m,k}, \theta^{i,p}_{m,k})$. The local measurement coordinates are used to update the object extent at all angles from the COO using a GP. The axis on the right side shows the change in the object radius as a function of angle from the COO.

This function is unknown and nonlinear and is mapped using a GP.

The GP maps the nonlinear function f in $r = f(\theta)$, shown in Fig. 2. The sensor measurements from scan to scan are considered as the training data for the GP regression. The multiple object posterior state estimation is intractable in the Bayesian framework. The estimation is achieved using the innovative convolutional kernels proposed in [15]. To improve the processing time, the measurement clustering is proposed as a pre-processing step. A novel object birth and death process based on the likelihood has also been proposed. The GPCPF requires the state sampling and the measurement sampling using the state and measurement models, respectively. The measurement samples are then updated with all the measurements based on the convolutional kernel. This is followed by the particle weight update and the state estimates. A re-sampling step is also proposed to cater for degeneracy.

3.1.3 Recent Advances

The GP based shape model has been proposed for multiple extended / group objects tracking in the random finite sets (RFS) framework in [20, 26], [27]. It has been shown that the performance of the algorithm improves due to better shape estimation with a GP model. The inference of GP shape model has also been proposed using a convolution particle filter (CPF) in [28] for a single object and in [15] for multiple objects using LiDAR data. The tracking snapshots are shown in Figure 3. Compared with the other complex shape models for EOT and GOT, the GPCPF based approach is not sensitive to the statistical properties of the object which is required by other models when measurements are coming from the object's surface. The GP based extended object tracking in 3D is demonstrated in [29]. A Bayesian object classifier based on a GP tracker is proposed in [30].

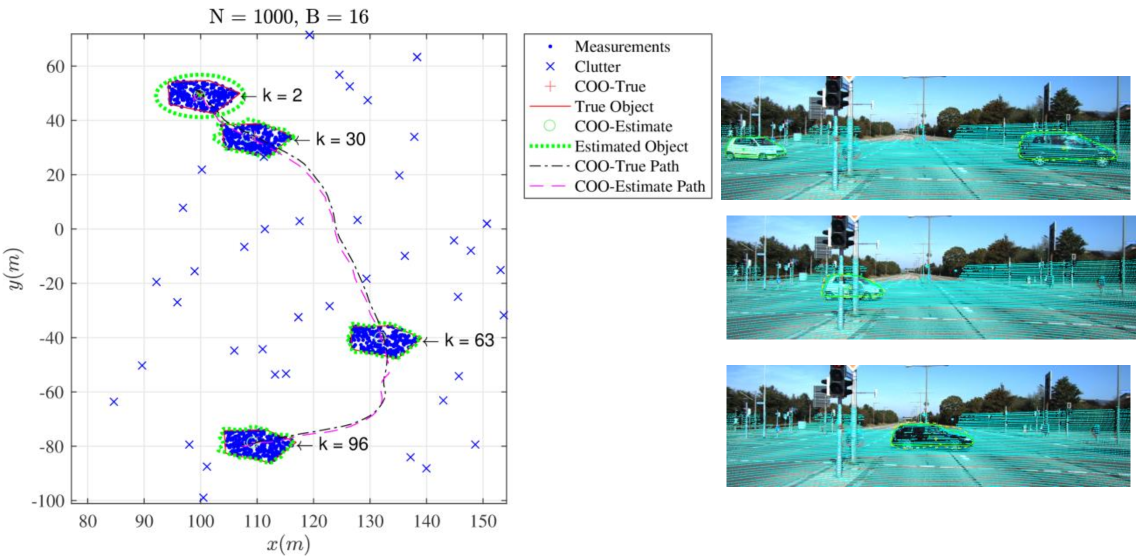

Figure 3: Tracking snapshots of single simulated object (left) and multiple real objects (right) using GPCPF.

The simulation consists of an irregularly shaped object moving through clutter (x). The true and estimated object COO, shape and trajectory is shown in the Figure. In the real data, the LiDAR measurements (moving and stationary objects) are used for tracking the moving objects (cars). The ground truth is created using the image data from a colour camera. The tracking results are overlaid on the corresponding camera images for reference.

4.0 PERFORMANCE EVALUATION

It is important to validate the developed EOT/GOT methods over benchmark datasets and performance metrics. Some of the important datasets and common metrics used in recent papers have been covered in this section.

4.1 Datasets

Although, the EOT and GOT applications are numerous, there is still a need of well systematised benchmark datasets and metrics that can characterise all aspects of the EOT and GOT process. Various publicly available datasets have been used for performance evaluation of the developed approaches. These data include PETS 2012, InLiDa [31], moving object thermal infrared imagery dataset (MOTIID) [32] and thermal infrared video (TIV) [33]. Although these and the datasets available for VOT consist of enough challenging scenarios, the annotation of the ground truth is not as per the EOT / GOT requirements. Hence, for spatial objects the ground truth data has to be manually created by the user in such cases. Two important datasets for pedestrian and group tracking are the DIAMOR [34] and ATC [35].

Standard datasets and evaluation metrics are also available for one of the fastest emerging EOT based application that is autonomous (ground) vehicle. Combinations of sensors, including LiDAR, are installed on the autonomous vehicle, which helps in the automatic navigation of the vehicle. These include the Ford campus vision and LiDar dataset [36], the Kitti vision benchmark suite [37], the Sydney urban objects [38], the Stanford track collection [39], the Oakland 3D point cloud dataset [40], and many more. Some useful surveillance camera datasets are also available such as ViSOR [41], VIRAT [42], EPFL [43] and CAVIAR [44].

4.2 Performance Metrics

The performance evaluation of extended / group object tracking needs further research. Although, various standard metrics for evaluation of multiple target tracking algorithms exist, with the optimal sub-pattern assignment metric (OSPA) [45] being the most commonly used metric nowadays, a standard metric for evaluation of the extended objects has not been proposed yet. Different metrics have been used to determine the performance some of which are given below:

- **Centre Distance** [28]**:** this is the distance (usually Euclidean) between the true and the estimated centre and its kinematics parameters. For multiple datasets or Monte Carlo based methods, the root mean square errors of the distance are more useful. If a represents the true and $\hat{a}$ represents the estimated parameter, then the Euclidean distance d is calculated as given below:

$$d = \sqrt{(a - \hat{a})^2} \,. \qquad (6)$$

- Orientation Distance [46]: this is the distance between the true and the estimated object orientation.
- Shape Precision and Recall: the shape precision and recall [47] have been used in computer vision for evaluating the shape detector and estimator performance. The shape recall evaluates the portion of true shape recalled and is related to the evaluation of the shape detection. The shape precision evaluates the portion of the estimator shape which is not part of the true shape and is related to the

  false detection. For a true shape $\boldsymbol{S}$ and an estimated shape $\hat{\boldsymbol{S}}$, the shape precision $\boldsymbol{P}$ and the recall $\boldsymbol{R}$ are given below:

$$P = \frac{S \cap \hat{S}}{\hat{S}}, \qquad R = \frac{S \cap \hat{S}}{S} \quad . \quad (7)$$

- Intersection over Union (IoU): also called the Jaccard similarity coefficient [48], combines the precision and recall in a single metric. The IoU is a ratio with value between 0 and 1, where 0 indicates a total mismatch and 1 indicates a total match. For a true shape S and an estimated shape $\hat{S}$, the IoU is calculated as given below:

$$IoU = \frac{S \cap \hat{S}}{S \cup \hat{S}} \quad . \qquad (8)$$

- Track life: is the time-span from the start to the end of the object's existence within the surveillance volume. It evaluates the track birth and death process. If k_s and k_e represent the start and the end time, respectively, then the track life T_l is given below:

$$T_l = k_e - k_s \qquad (9)$$

4.3 Additional Metrics for Multiple Objects

In case of multiple objects and clutter, additional metrics are included to evaluate especially the data association challenges. Some of these are the false alarms and track labelling errors. Although, these can be evaluated individually, research has also been done to propose a single metric that evaluates them together.

There are several different performance metric for methods that estimate object as an ellipse (ellipse being the most common shape model for applications now-a days) such as the Kulbeck-Lieber (KL) divergence, the Gaussian Wasserstein distance [46] etc. A modified Hausdorff distance is proposed for comparing two star convex shape models in [49]. For multiple point objects, the OSPA metric evaluates the localization and cardinality errors in a single metric. This is extended to evaluate the performance of tracking algorithms (where errors in track labelling are introduced) in [50]. The OSPA metric for comparing two extended object trackers has been proposed in [51], for elliptical shape models. The same metric has been modified for irregular shape trackers in [52] and is called modified OSPA (mOSPA). Other useful metrics are the multiple object tracking accuracy (MOTA) and multiple object tracking precision (MOTP) [53].

5.0 CONCLUSIONS AND FUTURE WORK

This paper presents recent advances of extended and group object tracking. Ttaxonomy of the main methods for tracking spatial objects with regular and irregular shapes is presented. These methods fuse multiple types of heterogeneous sensor data and deal with measurement origin uncertainties. Most current methods for tracking of spatial objects are model based. Current research is mainly focused on data driven, model free machine learning methods able to deal with data uncertainties and with large data volumes. Gaussian process methods are such machine learning methods, with a big potential. Results with recently developed Gaussian process methods are presented and their advantages over existing methods are demonstrated.

The spatio-temporal GP models and inference methods have a big potential of providing robust motion models for manoeuvring objects (both point and extended). Due to the data driven nature of the GP approaches, they can detect and model an infinite number of motion models. However, there are many theoretical questions to answer, including a thorough quantification of the propagation of the uncertainties and their impact on the final results. GP methods for high dimensional systems and data need cost efficient implementation, for real-time applications. Dealing with massive data is another challenge that is not yet resolved. Deep GP, convolutional GP methods can model complex motions and phenomena but the calculation of the on-line hyper-parameters still needs to be solved.

Another area, yet to be explored, is the application of OMGP for resolving data association problem arising in the EOT and GOT. The RFS based techniques have been primarily proposed for multiple EOT and GOT, but they are sensitive to the many model parameters. The data driven approach have the potential to provide solutions, which are less sensitive to unknown modelling parameters, deal with uncertainties and nonlinearities. They afford fusing different types of sensor data in an efficient manner. Other machine learning methods, such as convolutional neural networks and Bayesian neural networks have been mainly applied with imagery data.

Unmanned vehicles for air, land and sea are becoming popular in various commercial and military missions. This has brought a new dimension to the security, which includes the detection, tracking and identification of security hazard due to these types of vehicles. For low flying small air vehicles, LiDAR is being looked upon as one of the possible sensors [54]. The EOT and GOT methods become increasingly important for such autonomous systems.

Acknowledgements. The authors are grateful to the support SETA project funded by the European Unions Horizon 2020 research and innovation program under grant agreement no. 688082.

6.0 REFERENCES

[1] L. Mihaylova, A. Carmi, F. Septier and A. Gning, “Overview of Bayesian Sequential Monte Carlo Methods for Group and Extended Object Tracking,” *Digital Signal Processing,* vol. 25, pp. 1-16, February 2014.

[2] S. Rabinowitz, “Strategies for Automatic Track Initiation,” in *Advisory Group for Aerospace Research and Development,* NEUILLY-SUR-SEINE, 1979.

[3] K. Granström, M. Baum and S. Reuter, “Extended Object Tracking: Introduction, Overview, and Applications,” *Journal of Advances in Information Fusion,* vol. 12, no. 2, pp. 139-174, 2017.

[4] K. Gilholm, S. Godsill, S. Maskell and D. Salmond, “Poisson models for extended target and group tracking,” *Signal and Data Processing of Small Targets,* vol. 5913, p. 59130R, 2005.

[5] A. De Freitas, L. Mihaylova, A. Gning, D. Angelova and V. Kadirkamanathan, “Autonomous Crowds Tracking with Box Particle Filtering and Convolution Particle Filtering,” *Automatica,* vol. 69, pp. 380-394, 2016.

[6] K. Gilholm and D. Salmond, “Spatial Distribution Model for Tracking Extended Objects,” *IEE Proceedings-Radar, Sonar and Navigation,* vol. 152, no. 5, pp. 364-371, 2005.

[7] M. Feldmann, D. Franken and W. Koch, “Tracking of Extended Objects and Group Targets Using Random Matrices,” *IEEE Transactions on Signal Processing,* vol. 59, no. 4, pp. 1409-1420, 2011.

[8] A. Zea, F. Faion, M. Baum and U. D. Hanebeck, “Level-set Random Hypersurface Models for Tracking Nonconvex Extended Objects,” *IEEE Transactions on Aerospace and Electronic Systems,* vol. 52, no. 6, pp. 2990-3007, 2016.

[9] N. Wahlström and E. Özkan, “Extended Target Tracking using Gaussian Processes,” *IEEE Transactions on Signal Processing,* vol. 63, no. 16, pp. 4165-4178, 2015.

[10] J. l. Yang, P. Li and H.-w. Ge, “Extended Target Shape Estimation by Fitting B-spline Curve,” *Journal of Applied Mathematics,* pp. 1-9, 2014.

[11] B. Errasti-Alcala and P. Braca, “Track Before Detect Algorithm for Tracking Extended Targets Applied to Real-world Data of X-band Marine Radar,” in *Proceedings of the 17th International Conference on Information Fusion (FUSION),* 2014.

[12] P. Pileggi and M. Podt, “Improving Extent Estimation of Extended Targets with Track-before-Detect,” in *Proceedings of the 19th International Conference on Information Fusion (FUSION)*, 2016.

[13] R. Mahler, Advances in Statistical Multisource-Multitarget Information Fusion, Boston, USA: Artech House, 2014.

[14] A. De Freitas, L. Mihaylova, A. Gning, M. Schikora, M. Ulmke, D. Angelova and W. Koch, “A Box Particle Filter Method for Tracking Multiple Extended Objects,” in *IEEE Transactions on Aerospace and Electronic Systems*, 2018.

[15] W. Aftab, A. De Freitas, M. Arvaneh and L. Mihaylova, “A Gaussian Process Convolution Particle Filter for Multiple Extended Objects Tracking with Non-Regular Shapes,” in *Proceedings of the 21st International Conference on Information Fusion (FUSION)*, 2018.

[16] M. Wieneke and W. Koch, “A PMHT Approach for Extended Objects and Object Groups,” *IEEE Transactions on Aerospace and Electronic Systems*, vol. 48, no. 3, pp. 2349-2370, 2012.

[17] G. Vivone and P. Braca, “Joint Probabilistic Data Association Tracker for Extended Target Tracking Applied to X-band Marine Radar data,” *IEEE Journal of Oceanic Engineering*, vol. 41, no. 4, pp. 1007-1019, 2016.

[18] T. D. Vu and O. Aycard, “Laser-based Detection and Tracking Moving Objects using Data-driven Markov Chain Monte Carlo,” in *Proc. of the IEEE International Conference on Robotics and Automation*, 2009.

[19] K. Granström, C. Lundquist and O. Orguner, “Extended Target Tracking using a Gaussian-Mixture PHD Filter,” *IEEE Transactions on Aerospace and Electronic Systems*, vol. 48, no. 4, pp. 3268-3286, 2012.

[20] T. Hirscher, A. Scheel, S. Reuter and K. Dietmayer, “Multiple Extended Object Tracking using Gaussian Processes,” in *Proceedings of the 19th International Conference on Information Fusion (FUSION)*, 2016.

[21] K. Granström, L. Svensson, S. Reuter, Y. Xia and M. Fatemi, “Likelihood-based Data Association for Extended Object Tracking using Sampling Methods,” *IEEE Transactions on Intelligent Vehicles*, vol. 3, no. 1, pp. 30-45, 2018.

[22] P. J. Navarro, C. Fernandez, R. Borraz and D. Alonso, “A machine learning approach to pedestrian detection for autonomous vehicles using high-definition 3D range data,” *Sensors*, vol. 17, no. 1, p. 18, 2016.

[23] C. K. Williams and C. E. Rasmussen, “Gaussian Processes for Machine Learning,” *the MIT Press*, vol. 2, no. 3, p. 4, 2006.

[24] M. Lazaro-Gredilla, S. Van Vaerenbergh and N. D. Lawrence, “Overlapping Mixtures of Gaussian Processes for the Data Association Problem,” *Pattern Recognition*, vol. 45, no. 4, pp. 1386-1395, 2012.

[25] J. Hartikainen and S. Särkkä, “Kalman Filtering and Smoothing Solutions to Temporal Gaussian Process Regression Models,” in *Proceedings of the IEEE International Workshop on Machine Learning for Signal Processing (MLSP)*, 2010.

[26] R. P. Mahler, Advances in Statistical Multisource-Multitarget Information Fusion, Boston: Artech House, 2014.

[27] L. J. Chi, X. X. Feng and L. Miao, “Generalized Labeled Multi-Bernoulli Extended Target Tracking Based on Gaussian Process Regression,” in *Proceedings of the MATEC Web of Conferences*, 2018.

[28] W. Aftab, A. De Freitas, M. Arvaneh and L. Mihaylova, “A Gaussian Process Approach for Extended Object Tracking with Random Shapes and for Dealing with Intractable Likelihoods,” in *Proceedings of the 22nd International Conference on Digital Signal Processing (DSP)*, 2017.

[29] M. Kumru and E. Özkan, “3D Extended Object Tracking using Recursive Gaussian Processes,” in *Proceedings of the 21st International Conference on Information Fusion (FUSION)*, 2018.

[30] B. Tuncer, M. Kumru, E. Özkan and A. A. Alatan, “Extended Object Tracking and Shape Classification,” in *Proceedings of the 21st International Conference on Information Fusion (FUSION)*, 2018.

[31] C. Romero-González, Á. Villena, D. González-Medina, J. Martínez-Gómez, L. Rodríguez-Ruiz and I. García-Varea, “InLiDa: A 3D Lidar Dataset for People Detection and Tracking in Indoor Environments.,” in *Proceedings of VISIGRAPP (6: VISAPP)*, 2017.

[32] A. Akula, R. Ghosh, S. Kumar and H. Sardana, “Moving Target Detection in Thermal Infrared Imagery using Spatiotemporal Information,” *JOSA A*, vol. 30, no. 8, pp. 1492-1501, 2013.

[33] Z. Wu, N. Fuller, D. Theriault and M. Betke, “A Thermal Infrared Video Benchmark for Visual Analysis,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, 2014.

[34] F. Zanlungo, T. Ikeda and T. Kanda, “Potential for the Dynamics of Pedestrians in a Socially Interacting Group,” *Physical Review E*, vol. 89, no. 1, p. 012811, 2014.

[35] F. Zanlungo, D. Brščić and T. Kanda, “Spatial-size Scaling of Pedestrian Groups under Growing Density Conditions,” *Physical Review E*, vol. 91, no. 6, p. 062810, 2015.

[36] G. Pandey, J. R. McBride and R. M. Eustice, “Ford Campus Vision and Lidar Data Set,” *The International Journal of Robotics Research*, vol. 30, no. 13, pp. 1543-1552, 2011.

[37] A. Geiger, P. Lenz, C. Stiller and R. Urtasun, “Vision Meets Robotics: The KITTI Dataset,” *The International Journal of Robotics Research*, vol. 32, no. 11, pp. 1231-1237, 2013.

[38] M. De Deuge, A. Quadros, C. Hung and B. Douillard, “Unsupervised Feature Learning for Classification of Outdoor 3D Scans,” in *Proceedings of the Australasian Conference on Robitics and Automation*, 2013.

[39] A. Teichman, J. Levinson and S. Thrun, “Towards 3D Object Recognition via Classification of Arbitrary Object Tracks,” in *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA)*, 2011.

[40] D. Munoz, J. A. Bagnell, N. Vandapel and M. Hebert, “Contextual Classification with Functional Max-margin Markov Networks,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2009.

[41] R. Vezzani and R. Cucchiara, “Video Surveillance Online Repository (ViSOR): An Integrated Framework,” *Multimedia Tools and Applications*, vol. 50, no. 2, pp. 359-380, 2010.

[42] S. Oh, A. Hoogs, A. Perera, N. Cuntoor, C. C. Chen, J. T. Lee, S. Mukherjee, J. Aggarwal, H. Lee, L. Davis and others, “A Large-scale Benchmark Dataset for Event Recognition in Surveillance Video,” in *Proceedings of the IEEE conference on Computer vision and pattern recognition (CVPR)*, 2011.

[43] J. Berclaz, F. Fleuret, E. Turetken and P. Fua, “Multiple Object Tracking using K-shortest Paths Optimization,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 33, no. 9, pp. 1806-1819, 2011.

[44] O. Brdiczka, J. Maisonnasse and P. Reignier, “Automatic Detection of Interaction Groups,” in *Proceedings of the 7th International Conference on Multimodal Interfaces*, 2005.

[45] D. Schuhmacher, B. T. Vo and B. N. Vo, “A Consistent Metric for Performance Evaluation of Multi-object Filters,” *IEEE Transactions on Signal Processing*, vol. 56, no. 8, pp. 3447-3457, 2008.

[46] S. Yang, M. Baum and K. Granström, “Metrics for Performance Evaluation of Elliptic Extended Object Tracking Methods,” in *Proceedings of the IEEE International Conference on Multisensor Fusion and Integration for Intelligent Systems (MFI)*, 2016.

[47] C. Wolf and J. M. Jolion, “Object Count/Area Graphs for the Evaluation of Object Detection and Segmentation Algorithms,” *International Journal of Document Analysis and Recognition (IJDAR)*, vol. 8, no. 4, pp. 280-296, 2006.

[48] P. Jaccard, “Nouvelles Recherches sur la Distribution Florale,” *Bull. Soc. Vaud. Sci. Nat.*, vol. 44, pp. 223-270, 1908.

[49] L. Sun, S. Zhang, B. Ji and J. Pu, “Performance Evaluation for Shape Estimation of Extended Objects using a Modified Hausdorff Distance,” in *Proceedings of the IEEE International Conference on Information and Automation (ICIA)*, 2016.

[50] B. Ristic, B. N. Vo and D. Clark, “Performance Evaluation of Multi-target Tracking using the OSPA Metric,” in *Proceedings of the 13th Conference on Information Fusion (FUSION)*, 2010.

[51] C. Lundquist, K. Granström and U. Orguner, “An Extended Target CPHD Filter and a Gamma Gaussian Inverse Wishart Implementation,” *IEEE Journal on Selected Topics in Signal Processing*, vol. 7, no. 3, pp. 472-483, 2013.

[52] A. Daniyan, S. Lambotharan, A. Deligiannis, Y. Gong and W. H. Chen, “Bayesian Multiple Extended Target Tracking Using Labelled Random Finite Sets and Splines,” *IEEE Transactions on Signal Processing*, 2018.

[53] K. Bernardin and R. Stiefelhagen, “Evaluating Multiple Object Tracking Performance: The CLEAR MOT Metrics,” *Journal on Image and Video Processing*, p. 1, 2008.

[54] M. Hammer, M. Hebel, B. Borgmann, M. Laurenzis and M. Arens, “Potential of Lidar Sensors for the Detection of UAVs,” in *Proceedings of the Laser Radar Technology and Applications XXIII*, 2018.

[55] K. Gilholm, S. Godsill, S. Maskell and D. Salmond, “Poisson Models for Extended Target and Group Tracking,” *Signal and Data Processing of Small Targets*, vol. 5913, p. 59130R, 2005.

[56] K. Granström, L. Svensson, S. Reuter, Y. Xia and M. Fatemi, “Likelihood-based data association for extended object tracking using sampling methods,” *IEEE Transactions on Intelligent Vehicles*, vol. 3, no. 1, pp. 30-45, 2018.