DEGREE PROJECT IN THE FIELD OF TECHNOLOGY
ENGINEERING PHYSICS
AND THE MAIN FIELD OF STUDY
COMPUTER SCIENCE AND ENGINEERING,
SECOND CYCLE, 30 CREDITS
STOCKHOLM, SWEDEN 2017

Offline Sensor Fusion for Multitarget Tracking using Radar and Camera Detections

ANTON ANDERSSON

KTH ROYAL INSTITUTE OF TECHNOLOGY
SCHOOL OF COMPUTER SCIENCE AND COMMUNICATION

KTH

MASTER'S THESIS

Offline Sensor Fusion for Multitarget Tracking using Radar and Camera Detections

Anton Andersson
antonand@kth.se

Examiner: Patric Jensfelt
Supervisor at KTH: Örjan Ekeberg
Supervisors at Volvo: Johan Florbäck and Nasser Mohammadiha

June 3, 2017

School of Computer Science and Communication
KTH Royal Institute of Technology

Abstract

Autonomous driving systems are rapidly improving and may have the ability to change society in the coming decade. One important part of these systems is the interpretation of sensor information into trajectories of objects. In this master's thesis, we study an *energy minimisation* method with radar and camera measurements as inputs.

An *energy* is associated with the trajectories; this takes the measurements, the objects' dynamics and more factors into consideration. The trajectories are chosen to minimise this energy, using a gradient descent method. The lower the energy, the better the trajectories are expected to match the real world. The processing is performed *offline*, as opposed to in real time. Offline tracking can be used in the evaluation of the sensors' and the real time tracker's performance. Offline processing allows for the use of more computer power. It also gives the possibility to use data that was collected after the considered point in time.

A study of the parameters of the used energy minimisation method is presented, along with variations of the initial method. The results of the method is an improvement over the individual inputs, as well as over the real time processing used in the cars currently. In the parameter study it is shown which components of the energy function are improving the results.

Off-line sensorfusion för tracking av flera objekt med kamera och radardetektioner

Sammanfattning

Mycket resurser läggs på utveckling av självkörande bilsystem. Dessa kan komma att förändra samhället under det kommande decenniet. En viktig del av dessa system är behandling och tolkning av sensordata och skapande av banor för objekt i omgivningen. I detta examensarbete studeras en energiminimeringsmetod tillsammans med radar- och kameramätningar.

En *energi* beräknas för banorna. Denna tar mätningarna, objektets dynamik och fler faktorer i beaktande. Banorna väljs för att minimera denna energi med hjälp av gradientmetoden. Ju lägre energi, desto bättre förväntas banorna att matcha verkligheten. Bearbetning sker offline i motsats till i realtid; offline-bearbetning kan användas då prestandan för sensorer och realtidsbehandlingen utvärderas. Detta möjliggör användning av mer datorkraft och ger möjlighet att använda data som samlats in efter den aktuella tidpunkten.

En studie av de ingående parametrarna i den använda energiminimeringsmetoden presenteras, tillsammans med justeringar av den ursprungliga metoden. Metoden ger ett förbättrat resultat jämfört med de enskilda sensormätningarna, och även jämfört med den realtidsmetod som används i bilarna för närvarande. I parameterstudien visas vilka komponenter i energifunktionen som förbättrar metodens prestanda.

Contents

Abstract **iii**

v

1 Introduction **1**
1.1 Current state of autonomous driving systems 1
1.2 Purpose 2
1.3 Problem formulation 3
1.4 Limitations 3

2 Background **5**
2.1 Multi-target tracking 5
2.2 Sensor fusion 6
2.3 Data association 6
2.4 Kalman filters and particle filters 6
2.5 Energy based methods 7
2.6 Comparisons 8

3 Method **9**
3.1 Available sensor data 9
3.2 Energy minimisation 10
3.3 Metrics for evaluation 16
3.4 Adaptations and improvements 18

4 Results **23**
4.1 Measurement errors 23
4.2 Parameter study 24
4.3 MOTP - Energy correlation 26
4.4 Method evaluation 27

5 Discussion **31**
5.1 Measurement errors 31
5.2 Parameter study 32

5.3 MOTP - Energy correlation 33
5.4 Method evaluation 33
5.5 Comments 34
5.6 Future work 35
5.7 Conclusions 35

Bibliography **37**

List of Symbols

$b(\mathbf{X}_i^t)$	shortest distance to the border of the field of view
c_t	number of matches at time frame t
$\mathbf{D}_g^t$	coordinates of detection g in time frame t
e_i	last frame for which the object i is defined
$e \approx 2.178$	Euler's number
$E(\mathbf{X})$	energy function that should be minimised
E_{app}	appearance component of the energy
E_{det}	detection component of the energy
E_{dyn}	dynamic component of the energy
E_{exc}	exclusion component of the energy
E_{reg}	regularisation component of the energy
E_{per}	persistence component of the energy
fp_t	number of false positives in time frame t
F	number of time frames in the sequence
$F(i)$	number of frames of object i
ϕ_i	angular coordinate of object i
g_t	number of ground truth objects in time frame t
λ	penalty given to each detection in the detection energy
mme_t	number of mismatches in time frame t
m_t	number of misses in time frame t
M	set of correspondences between hypotheses and ground truth objects
N	number of found objects
$\mathcal{O}_{ij}$	occlusion matrix, the fraction of object i that is covered by object j
R_i	distance to object i
s	world frame size of the objects
s_i	first time frame for which the object i is defined
$\tilde{s}_i$	size of object i in the image plane
T	threshold for a hit/miss
v_i^t	visibility of object i in time frame t
V_{ij}	relative overlap of object i and object j
w_g^t	confidence of detection g in time frame t

$\mathbf{X}$	state vector containing coordinates of all objects in all time frames
$\mathbf{X}_i^t$	coordinates of object i in time frame t

Chapter 1

Introduction

This thesis is made in collaboration with Volvo Cars. In the introduction, the current state of autonomous vehicle research is presented along with the problem formulation for this thesis.

1.1 Current state of autonomous driving systems

Modern cars are equipped with many different kinds of sensors used to add awareness of the surroundings. Revolution-counters have for a long time been used to measure the cars' own speed and ultrasonic parking sensors have later become common. With a shift to active safety systems new kinds of sensors such as radars and cameras are appearing. Cameras can be used to identify and estimate the position of outside objects. Radars can measure the distance to objects and are robust to dust and and adverse weather [22]. By analysing the doppler shift of the returning signal, a radar can give the relative speed of other objects in addition to their relative position.

In autonomous cars, even more sensors are needed in order to fully understand the surrounding. Cameras and radars , as well as lidars, are incorporated in these cars. A lidar illuminates the surrounding objects with laser light and measures the reflected light. Lidars perform outstandingly well in terms of its angular resolution and range measurements but are sensitive to fog, rain and snow [18].

During the recent years, autonomous driving and active safety systems have been hot fields for research and development. Automatic braking

with systems that can see other cars and pedestrians can be found on production cars. These systems often come with limitations such as inabilities to recognise objects in certain situations like a person bending over or carrying a large box [12]. Many car manufacturers as well as tech companies such as Google and Uber are performing tests with fully autonomous vehicles in real traffic environments. Google has currently driven more than 2.4 million kilometers autonomously [17]. Autonomous small buses have recently been tested with passengers in Helsinki, although only at speeds up to 11 km per hour and in easy environments [20]. Currently produced Tesla cars have self-driving abilities that rely on the data from radars and cameras, while for example Google also use a lidar sensor [21].

There are many ethical aspects to consider in the development and deployment of autonomous vehicles. These include deciding when the systems are good enough to be deployed and who should be responsible for accidents.

Volvo Cars has historically been and is currently considered to be on the forefront of vehicle safety. To further increase the safety of its cars, surrounding vehicles and pedestrians, Volvo Cars is heavily investing in safety research. The focus has moved from mostly trying to limit the collateral damage of a crash to preventing crashes with dynamic driving controls. Dynamic driving controls, such as Electronic Stability Control, are second only to seat belts in terms of increasing road safety [2].

1.2 Purpose

The accuracy and reliability of active safety systems used in modern cars need to be verified to ensure proper functionality. This is true for current active safety systems and is increasingly important for the up-and-coming development of autonomous vehicles. When verifying the systems, a larger amount of computational power is accessible compared to the in-car computer. Information from the entire scenario is also available; in the real-time implementation only information gathered before a perticular moment is at hand. It is therefore of interest to use different tracking approaches for the offline tracking used in verification and the real-time tracking used in the cars.

1.3 Problem formulation

In this thesis we examine an energy minimisation method for tracking, described by Milan et al. [14]. The main objective of this thesis is to investigate how well this energy minimisation method performs at offline tracking with in-car sensors in a driving environment.

The questions we seek to answer are: Does the energy used in the method correspond to how close the trajectory is to reality? Can availability, false positives rate, and position accuracy of traffic objects obtained by real time algorithms be improved using the energy minimisation method? How can this be done in a good way and how much of an improvement can be achieved?

1.4 Limitations

In this thesis, only the energy minimisation method proposed by Milan et al. will be considered and compared with the currently used real-time system. We can therefore not draw conclusions about how well it works compared to other methods.

It is possible to access raw camera images; however, in this thesis only precalculated positions of obstacles will be used. This means that we do not consider shape and color of objects in the tracking.

Different types of objects have different forms and dynamic characteristics. We limit ourselves to tracking other cars.

Chapter 2

Background

In the background chapter, the problem is presented along with contextual background. We also present different approaches to the problem including the method used in the thesis.

2.1 Multi-target tracking

The aim of multi-target tracking is to estimate the trajectories of several moving objects. This has military and civil uses such as tracking missiles and analysing crowd behaviour when evacuating buildings. Tracking often involves data association; resolving correspondence between trajectories and sensor measurements. This problem quickly leads to difficult combinatorial problems when many trajectories are possible matches. Trajectory management is also needed for filtering out noise from measurements. In some implementations multiple sensors are used, calling for sensor fusion to improve the precision of the trajectories [14].

Aspects to consider when designing a tracker include: [11]

- The number of tracked objects may vary.
- The objects may enter and leave the sensors' field of view.
- Objects may temporarily occlude each other.
- False reflections and noise may give false readings from the sensors.

2.2 Sensor fusion

In sensor fusion we consider the problem of having one or many objects being observed by several different sensors. Each of these sensors have different errors in measurement and noise characteristics. The goal of the sensor fusion is to obtain a state-vector describing the position of the tracked objects that is better, in terms of accuracy and reliability, than the individual outputs of the sensors [6].

Sensor fusion methods are generally divided into three categories; centralised, decentralised and hybrid sensor fusion. In centralised sensor fusion, the sensor readings are brought together without preprocessing before trajectory tracking begins. This requires a large bandwidth for transmission. In the other extreme is decentralised fusion. The individual sensors are preprocessed in terms of noise removal and trajectory assignment and are then fused at a higher level. In centralised sensor fusion, it is possible to make more precise predictions. The system is also more difficult to modify since all changes have to be made in the often complicated and fine-tuned central unit [13].

2.3 Data association

Some approaches to sensor fusion require that each sensor measurement is associated with the ID of an object. The data association problem for multiple target tracking can be solved with several different techniques, such as Global Nearest Neighbour which ensures that the global statistical distance is minimised. Another technique is Suboptimal Nearest Neighbour, which sacrifices accuracy for speed. They all have in common that the input is a list of sensor observation and the output is a matching between these observations and objects [9].

2.4 Kalman filters and particle filters

There are many approaches to solving the multisensor data fusion problem, of which Kalman filtering has been the most widely used [6]. The process includes a first stage of making a prediction of the state using a

linear physical model and the previous state. In the second stage this prediction is combined with the next measurement using a weighted average to get the next state estimate [8]. Kalman filters return the optimal estimate in a least squares sense assuming that the underlying dynamics and measurement models are linear and that the errors are Gaussian [14]. The Kalman filter can be applied on each sensor before combining the filtered sensor data, or one can use a single Kalman filter on fused sensor measurement. The estimations from the former approach have a lower computational cost and can be performed in parallell while the latter approach generally produces better results [6].

There are other methods that handle models that are not linear. One such method is the extended Kalman filter (EKF). A linear model is estimated with a first order Taylor expansion. The accuracy of the EKF is dependent on the estimation error being sufficiently small, which is hard to guarantee. Bad estimations can lead to instability [5].

In particle filters the state vector is represented by a set of random samples which are recursively updated and compared to the measurements. The very unlikely samples, those that do not fit the measurements, are discarded and new samples are created around the likely ones. This representation of the state vector enables the algorithm to handle non-Gaussian state probability density functions [7].

Both Kalman filters and particle filters are originally made for single target tracking. In order to handle multiple targets the objects are handled independently with an initial data association which divides the input sensor data into sets for each target [14].

2.5 Energy based methods

One popular approach to the multiple object tracking problem is to formulate a heuristic function (called an *energy function*) that takes multiple aspects of the scene into account. This energy function maps each possible solution to an energy. By minimising the energy function, all trajectories are calculated on a global level in a single process [14].

One energy minimisation approach is proposed by Rodriquez et. al [19]. A crowded scene is modelled both with a crowd density estimation that

does not take individuals' locations into account and with a finer detailed position estimation of individuals. The crowd density estimation is a way to deal with the difficulties of tracking individuals in a crowd. An energy function is formulated that punishes solutions where the individual movements do not match the crowd density model. This energy function is then minimised using a greedy search procedure [19].

The energy minimisation method that we focus on in this thesis was first proposed by Andriyenko et al. [1] and improved upon by Milan et al. [14]. An energy function is formulated that represents the problem as faithfully as possible while still being differentiable. The energy function includes several components including one that is limiting the motion of the tracked targets, one that limits the relative position of the targets and one that limits where new targets are allowed to appear or disappear.

2.6 Comparisons

In the study by Nguyen et al. [16], a comparison between different tracking approaches is presented. Methods proposed by Andriyenko et al. [1], Milan et al. [15], Yan et al. [23], Chau et al. [4] and Nguyen et al. [16] are compared for the TUD-Stadtmitte dataset which contains challenges such as highly crowded environments with frequent occlusions and many similar objects.

The approach proposed by Milan et al. fares best or matches the performance of the other trackers for all metrics used. Using the MOTP evaluation metric which measures the ability of the tracker to estimate the object position [3], Milan et al.'s approach matches the Nguyen et al. approach while outperforming Andriyenko et al.'s with a score of 0.65 compared to 0.63. Milan et al. also matches the performance of Nguyen et al. and Yen et al. using the MT metric which measures the ratio of objects that are mostly tracked. Using the MOTA metric which measures the amount of misses, false positives and mismatches [3], Milan et al. is the best performing tracker followed by Andriyenko et al. and Nguyen et al. [16].

Chapter 3

Method

In this chapter, the used minimisation method along with performance metrics are described. The method is largely adapted from the work of Milan et al. In the end of the chapter adaptations and improvements are presented.

3.1 Available sensor data

In this thesis we will use Volvo Cars' off-the-shelf camera, radar and lidar sensors. A wide coverage is provided with a mid range radar and a more narrow field of view with a long range radar.

Due to how the camera and radar operates, we expect the camera to have a smaller azimuthal error and the radar to have a smaller error in distance. This is investigated by, for each object position from the camera and radar, we find the closest point from the lidar (ground truth) and measure the distance. A threshold for a maximum deviation between the two points is used to filter out cases where there is no match.

A lidar attached to the roof of the car is used as ground truth. It has a 360 degree view of the surrounding and can see targets up to 174 meters away [10].

We compare our results with a tracker that is now used to interpret the sensor measurements in real time for test cars. We do not know what method is implemented in this tracker but since this is the method that is used currently, we find it relevant to compare our results to it. The relative performance of this tracker is mentioned with the results but the exact values are not shown in the graphs.

For this thesis, we use approximately one hour of recorded sensor information. The provided data was recorded with Volvo Cars' test vehicles in a real world suburban traffic environment. Some pre-processing has been performed inside the sensors, likely with a Kalman filter.

The camera and radar operate at different update frequencies. In order to simplify the implementation, the camera detections are in our work interpolated to match the time points of the radar.

3.2 Energy minimisation

Energy minimisation methods have become popular for solving the sensor fusion problem with multiple targets [14]. Andriyenko et al. [1] propose such a method which Milan et al. [14] improves upon. This method has been proven to fare well when compared to other methods (see section 2.6). The multi-target tracking problem is formulated as a minimisation of an "energy" which is calculated from properties of the trajectories. This energy is defined in continuous space and takes into account how well the tracks align with the measurements, if the tracks comply with the physical constrains of the objects in terms of acceleration, occlusion of the objects and trajectory persistence. The energy is dependent on all targets in all frames; target positions are estimated also when they are occluded. Due to the complex nature of the energy function, it is unlikely to be convex and the solution is thus not guaranteed to be globally optimal. In order to find good local minimas, the conjugate gradient descent method is used. Since the energy function is defined in closed form we can perform the gradient descent in a computationally efficient manner. Trans-dimensional jump moves, i.e. moves that change the number of trajectory points, can be performed in order to decrease the importance of a good initialisation.

Symbol	Description
$\mathbf{X}$	State vector containing coordinates of all objects in all frames
$\mathbf{X}_i^t$	The world coordinates of object i in time frame t
s_i	The first frame for which the trajectory i is defined
e_i	The last frame for which the trajectory i is defined
N	The number of found objects
F	The number of time frames in the sequence
$\mathbf{X}_i^t$	The world coordinates of object i in time frame t
$E(\mathbf{X})$	The energy function that should be minimised
v_i^t	The visibility of object i in time frame t
$\mathbf{D}_g^t$	The coordinates of detection g in frame t
w_g^t	The confidence of detection g in frame t
s	The world frame size of the objects
$\tilde{s}_i$	The size of object i in the image plane

The goal of the method described by Milan et. al is to find a set of trajectories for the N found objects in the F frames of our sequence. The state vector $\mathbf{X}$ contains the coordinates of all objects found in all frames. $\mathbf{X_i^t}$ is a duplet (x,y) with the location of target i in frame t. The location of target i is defined for all frames between s_i and e_i. The energy function $E(X)$ associates a cost with our state vector $\mathbf{X}$. We solve the tracking problem by finding the $\mathbf{X}$ that minimises:

$$\mathbf{X}^* = \arg\min_{\mathbf{X}} \mathbf{E}(\mathbf{X}) \tag{3.1}$$

The energy function used consists of six components:

- E_{det} accounts for the trajectories' distance to observations.
- E_{app} keeps the object appearance consistent in all parts of its trajectory.
- E_{dyn} constraints the objects acceleration.
- E_{exc} keeps two or more objects from occupying the same physical space.
- E_{per} prevents objects from appearing or disappearing inside of the tracked area.

- E_{reg} keeps the solution simple and prevents overfitting by favouring few trajectories.

The energy function is a linear combination of these terms:

$$E = E_{det} + \alpha E_{app} + \beta E_{dyn} + \gamma E_{exc} + \delta E_{per} + \epsilon E_{reg} \tag{3.2}$$

The effects of the detection, dynamic and persistence components of the energy can be seen in figure 3.1.

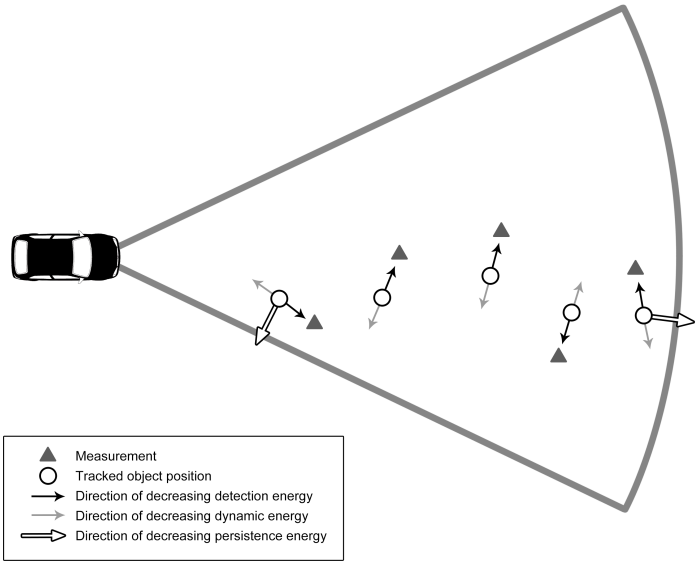


FIGURE 3.1: A visual representations of the detection, dynamic and persistence components of the energy

The observational model part of the energy, E_{det}, increases smoothly when the distance from the estimated position $\mathbf{X}_i^t$ to the detection $\mathbf{D}_g^t$ increases. The size s of the target is taken into account; a greater deviation between the estimated position and the detection is tolerated when s is large. In order to take occlusion into account the visibility v_i^t of each target is calculated. Each target is given a penalty $v_i^t \cdot \lambda$; targets that can not be seen by the sensors are not penalised for being far from the detections.

$$E_{det} = \sum_{i=1}^{N} \sum_{t=s_i}^{e_i} \left[v_i^t \cdot \lambda - \sum_{g=1}^{D(t)} \omega_g^t \frac{s^2}{\|\mathbf{X}_i^t - \mathbf{D}_g^t\|^2 + s^2} \right] \tag{3.3}$$

The occlusion handling is performed with a closed form, differentiable formulation. The bounding box indicator function, i.e. the area of the field of view that is occupied by a target, is approximated as a anisotropic Gaussian

$$N_i(\tilde{\mathbf{x}}) := N(\tilde{\mathbf{x}}_i, C_i) \tag{3.4}$$

where $\tilde{\mathbf{x}}_i$ is the image coordinates of target i and the variance

$$\mathbf{C}_i = \begin{pmatrix} \frac{1}{2}(\frac{\tilde{s}_i}{2})^2 & 0 \\ 0 & (\frac{\tilde{s}_i}{2})^2 \end{pmatrix} \tag{3.5}$$

with $\tilde{s}$ being the height of the target in the image plane. Milan et al. assumes all targets to be humans; the factor $\frac{1}{2}$ makes the Gaussian approximately take the shape of a human. The relative overlap V_{ij}, i.e. the fraction of overlap between two objects is had by integrating the overlap functions over the image plane. This can be computed by

$$V_{ij} = \exp(-\frac{1}{2}[\mathbf{x}_i - \mathbf{x}_j]^\top \mathbf{C}_{ij}^{-1}[\mathbf{x}_i - \mathbf{x}_j]) \tag{3.6}$$

with $\mathbf{C}_{ij} = \mathbf{C}_i + \mathbf{C}_j$. By symmetry we have $V_{ij} = V_{ji}$.

In order to take the relative depth of the objects into account in a closed form differentiable way, a sigmoid along the vertical axis (y-axis) centered on y_i is used. This assumes that the targets' order in the y-axis is the same as the depth order of the objects.

$$\sigma_{ij} = \frac{1}{1 + e^{y_i - y_j}} \tag{3.7}$$

An occlusion matrix, $\mathcal{O}_{ij}$, is defined with $\mathcal{O}_{ij} = \sigma_{ij} \cdot V_{ij}, i \neq j$ and $\mathcal{O}_{ii} = 0$. Element $\mathcal{O}_{ij}$ corresponds to the fraction of object i that is covered by object j in the image plane. The total fraction of object i that is occluded is $\sum_j \mathcal{O}_{ij}$, assuming that no part of the object is occluded by two other objects. The visibility v_i is given by

$$v_i = e^{-\sum_j \mathcal{O}_{ij}} \tag{3.8}$$

The appearance part of the energy, E_{app}, keeps the object appearance consistent in all parts of its trajectory. Milan et al. use RGB color histograms within the object target box as defined in equation (3.4) for this.

The dynamic part of the energy, E_{dyn} constraints the objects acceleration. This reduces the number of identity switches (which often result in a different velocity both in magnitude and direction) and it smoothes errors in the detections. The velocity for object i at time t and at time $t+1$ is proportional to $\mathbf{X}_i^{t+1} - \mathbf{X}_i^t$ and $\mathbf{X}_i^{t+2} - \mathbf{X}_i^{t+1}$ respectively. The acceleration is proportional to

$$\mathbf{X}_i^{t+2} - \mathbf{X}_i^{t+1} - (\mathbf{X}_i^{t+1} - \mathbf{X}_i^t) = \mathbf{X}_i^t - 2\mathbf{X}_i^{t+1} + \mathbf{X}_i^{t+2} \tag{3.9}$$

The dynamic energy is calculated by taking the sum of the consecutive changes in the velocity vector

$$E_{dyn} = \sum_{i=1}^{N} \sum_{t=s_i}^{e_i-2} \left\| \mathbf{X}_i^t - 2\mathbf{X}_i^{t+1} + \mathbf{X}_i^{t+2} \right\| \tag{3.10}$$

The mutual exclusion part of the energy, E_{exc} keeps two or more objects from occupying the same physical space. It also keeps objects far enough apart to keep one observation from resulting in more than one tracked object. The energy of two overlapping obstacles goes to infinity as their distance approaches zero

$$E_{exc} = \sum_{t=1}^{F} \sum_{i,j \neq i}^{N(t)} \frac{s}{\|\mathbf{X}_i^t - \mathbf{X}_j^t\|^2} \tag{3.11}$$

Trajectories starting or ending in the interior of the tracked area are penalised. A sigmoid along the edges of the tracked is used to do this:

$$E_{per} = \sum_{i=1, t \in \{s_i, e_i\}}^{N} \frac{1}{1 + \exp(-q \cdot b(\mathbf{X}_i^t) + 1)} \tag{3.12}$$

with $b(\mathbf{X}_i^t)$ being the shortest distance to the border of the area being observed and $q = \frac{1}{s}$ determines how fast the penalty decreases along the border.

Finally, a regulating term is used to keep the the solution simple and prevent overfitting. This consists of both a term for keeping the total number of targets low and a term that penalises short tracks over long

$$E_{reg} = N + \sum_{i=1}^{N} \frac{1}{F(i)} \tag{3.13}$$

A standard conjugate gradient descent method is used to minimise the energy function. Due to the non-convex nature of the energy function, the choice of initialisation is important. Milan et al. uses the output of an per-target extended Kalman filter as the initial value.

In order to weaken the dependency on a good initialisation, jump moves are performed upon convergence of the gradient descent. These jump moves may change the dimensionality of the state vector, i.e. they may add or remove points to the trajectories or change the number of targets. Several different types of jump moves are executed: growing, shrinking, merging, splitting, adding and removing of trajectories. These are iterated through as described in Alg.1 (adapted from [15]).

The growing and shrinking of the trajectories is performed to pick up lost parts of trajectories and get rid of false ones. This is achieved by adding t positions to the front, $\tilde{\mathbf{X}}_i = (\mathbf{X}_i^{s_i-t:s_i-1}, \mathbf{X}_i)$, or the back $\tilde{\mathbf{X}}_i = (\mathbf{X}_i, \mathbf{X}_i^{e_i+1:e_i+t})$. The shrinking is performed either on the front $\tilde{\mathbf{X}}_i = \mathbf{X}_i^{s_i+t:e_i}$, or the back $\tilde{\mathbf{X}}_i = \mathbf{X}_i^{s_i:e_i-t}$.

The merging and splitting is performed to remove instances of one real world object switching ID or two objects exchanging ID. This is more likely to happen when the objects are occluded. Merging, $\tilde{\mathbf{X}}_k = (\mathbf{X}_i, \mathbf{X}_{\text{con}}, \mathbf{X}_j)$, can be performed when two paths can be connected while preserving plausible motion.

Adding is performed where there are strong detections that are not yet a part of any trajectory and a trajectory is removed if its contribution to the energy is above a certain limit. An added trajectory initially exists in three

Algorithm 1 Energy minimisation

Input: K initial solutions, detections D
Output: Best of $\leq K$ solutions
  for $k = 1$ to K **do**
    while $\neg$ converged **do**
      for $m \in$ {grow,shrink, add, remove, merge, split} **do**
        for $i \in 1, ..., N$ **do**
          try jump move m on trajectory i
          if $E^{new}(\mathbf{X}_k) < E^{old}(\mathbf{X}_k)$ **then**
            perform jump move
          end if
        end for
        perform conjugate gradient descent
      end for
    end while
  end for
  return arg $\min_{\mathbf{X}_k} E(\mathbf{X}_k)$

frames $\tilde{\mathbf{X}}_i^{t-1:t+1} = (\mathbf{D}_g^t, \mathbf{D}_g^t, \mathbf{D}_g^t)$, that later may be extended or merged with other trajectories.

3.3 Metrics for evaluation

Common metrics are needed for assessment and comparisons of the performance of different trackers. The metrics that we use should be able to evaluate trackers ability to:

- give accurate estimations. The estimated positions of the tracked objects should be close to the real positions.
- find all objects. The tracker should not miss objects.
- not find objects where there are no objects. The tracker should not give false positives.
- keep track of individual objects. The ID of an object should not change in the middle of a path, for example when the object is occluded.

We also want the metrics to be straight forward; they should have few free parameters and behave according to human intuition [3].

One metric that has been widely accepted by the tracking community is the CLEAR MOT evaluation (Bernardin and Stiefelhagen, [3]) [14]. Bernardin and Stiefelhagen propose two metrics: the multiple object tracking accuracy (MOTA) and the multiple object tracking precision (MOTP). These metrics are suitable for general performance evaluation. They take into account the tracker's precision in terms of estimating the obstacle position, its ability to not miss obstacles and how it is keeping consistent tracks over time [3].

Following Bernardin and Stiefelhagen, for each time frame t we establish the best correspondence between the the hypothesis h_j and the ground truth object o_i. We then compute the error in the position estimation for each correspondence, d_t^i which gives us the *multiple object tracking precision* (MOTP):

$$\text{MOTP} = \frac{\sum_{i,t} d_t^i}{\sum_t c_t} \tag{3.14}$$

with c_t being the number of matches at time t.

In order to get the *tracking accuracy* we compute the correspondence errors:

- The ground truth objects that have no corresponding hypothesis are regarded as *misses*.
- The hypotheses that have no corresponding ground truth object are *false positives*.
- Occurrences when the tracking hypotheses is changed from frame $t-1$ to frame t are *mismatch errors*. Two object that switch ID or objects that change ID after having been occluded are examples of this.

In order to find the correspondence between the hypotheses and the ground truth objects we start by setting a maximum for how far apart we allow the hypothesis to be from the ground truth object. If the distance is more than a certain threshold T we consider the object to be missed. The value for T is determined by the size of the tracked objects. We start with no correspondences, $M_0 = \{\cdot\}$. If there already is a matching between o_i and h_j in

the previous time frame and the distance still is less than T, a correspondence is made between o_i and h_j and this is added to M. The mapping between the hypothesises and ground truth objects for which there is no matching hypothesis is done in a way that minimises the sum of the distance errors.

In order to get the number of mismatches, we consider the set of matchings, $M_t = \{(o_i, h_j)\}$, made up until time t. If we at time $t + 1$ get a contradicting matching between o_i and h_k, we increment the number of mismatches, mme_t and we replace (o_i, h_j) with (o_i, h_k) in M_{t+1}.

The hypotheses for which there is no matching in M to a ground truth object are counted as false positives. fp_t is the number of false positives. All ground truth objects for which there is no matching hypothesis are counted as misses and m_t is the number of misses.

The multiple object tracking accuracy (MOTA) is given by

$$\text{MOTA} = 1 - \frac{\sum_t (m_t + fp_t + mme_t)}{\sum_t g_t} \tag{3.15}$$

where g_t is the number of ground truth objects in frame t. More misses, false positives and mismatches results a lower score and a perfect tracker will get a score of 1.

3.4 Adaptations and improvements

Milan et al. uses data from a camera located above the targets pointing at an angle downwards. The radar and camera used in this thesis are located at the same level as the targets and are pointing forwards (see figure 3.2). This simplifies the occlusion handling; polar coordinates are used and the indicator function is approximated as a one-dimensional Gaussian in the angular coordinate. The camera and radars are in our case moving with the car, so the position of the car has to be taken into consideration in the occlusion handling.

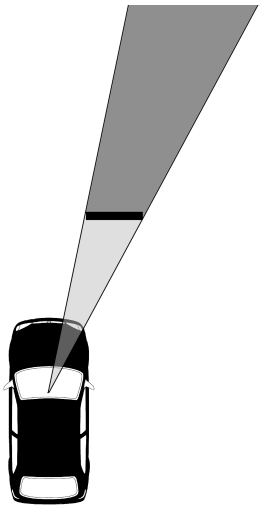


FIGURE 3.2: The occluded area (dark grey) behind an object.

We use the angular size of the target

$$\theta_i = \tan^{-1}(\frac{s}{R_i}) \tag{3.16}$$

with R_i being the distance to the target i. The bounding box indicator function is a normal distribution around the angular coordinate, ϕ_i, of the target

$$N_i(\tilde{\mathbf{x}}) := N(\phi_i, C_i) \tag{3.17}$$

and the variance

$$C_i = \frac{1}{2}\left(\frac{\theta_i}{2}\right)^2 \tag{3.18}$$

The relative overlap is

$$V_{ij} = \exp(-\frac{1}{2}|\phi_i - \phi_j|^2 C_{ij}) \tag{3.19}$$

with $C_{ij} = C_i + C_j$. We are taking the depth ordering of targets into account with a sigmoid along the radial dimension

$$\sigma_{ij} = \frac{1}{1 + e^{R_i - R_j}} \tag{3.20}$$

An occlusion matrix, $\mathcal{O}_{ij}$, is defined with $\mathcal{O}_{ij} = \sigma_{ij} \cdot V_{ij}$. And finally we get the visibility v_i

$$v_i = e^{-\sum_j \mathcal{O}_{ij}} \tag{3.21}$$

The radar is operating at approximately twice the frame rate as the camera. In order to simplify implementation, the camera detections were interpolated to match the same points in time as the radar.

In the scenario that Milan et al. is considering, the field of view is a static rectangle for the persistence part of the energy. In our case the field of view is more complicated, with the long range radar and mid-range radar having different fields of view (figure 3.3). Since the car is moving, we calculate the persistence term from positions in local coordinates.

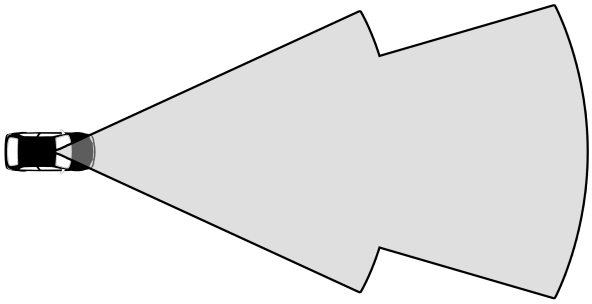


FIGURE 3.3: The field of view for the long and mid-range radar combined. Not to scale.

The raw image is not used; the appearance model could use the radar cross-section instead of the colour histogram, but this would require an association between detections and trajectories. The appearance model is therefore left out of our implementation.

Milan et al. uses only data from a camera. Since we use both radar and camera detections, it is of interest to take the reliability of these sensors into account in the observational part of the energy. This is performed by scaling the distance from trajectory to detections with the standard deviation of the error of the respective sensors in the longitudinal and latitudinal directions.

$$E_{det} = \sum_{i=1}^{N} \sum_{t=s_i}^{e_i} \left[v_i^t \cdot \lambda - \sum_{g=1}^{D(t)} \frac{s^2}{\|\boldsymbol{\omega}_g \cdot (\mathbf{X}_i^t - \mathbf{D}_g^t)\|^2 + s^2} \right] \tag{3.22}$$

ω_g is chosen as the standard deviation of error of the sensor type of in the longitudinal and latitudinal direction.

Measurements from both radar, camera and lidar are used. The mid-range radar trajectories are used as initialisation. In order for us to be able to evaluate our results, we need lidar measurements for the tracked objects. We are not interested in estimating the position of static objects. Before applying the energy minimisation method, we pre-process the data. This is done by only considering radar measurements that are within a certain distance (5 meters) from a lidar track. We also only consider lidar tracks that are have a length of more than 30 meters. This both removes static objects and lets us focus on longer tracks. Different data sets were used in the parameter search and for the evaluation. In the parameter search, four minutes of driving data was used with about 60 tracked vehicles. For the evaluation, ten minutes of driving data was used with about 300 tracked vehicles.

The ground truth that we use in this project is lidar readings. These are not complete in the sense that some real world objects may not be sensed by the lidar. However, we assume that all detections by the lidar are corresponding to real world objects. We redefine *false positives*: a hypothesis is a false positive only if it has been or will be matched to a lidar observation but currently is not. The lidar detection trajectory also has to exist at the current time frame (see figure 3.4).

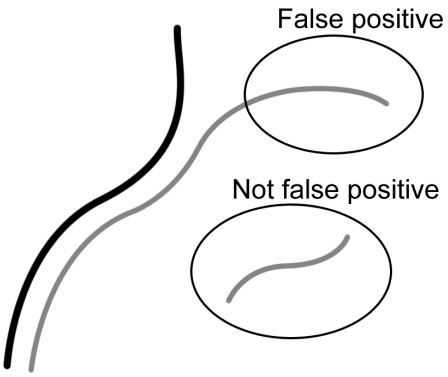


FIGURE 3.4: Ground truth (black line) and hypothesis (grey line). A visualisation of what is considered a *false positive*.

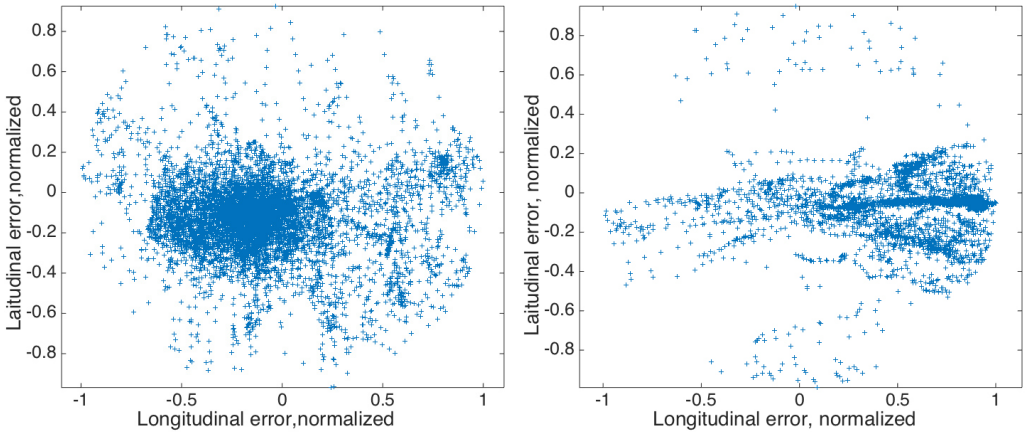
Chapter 4

Results

In the results chapter, the results generated by the method are presented. These are compared to the individual sensors and the currently used real-time tracking. We also demonstrate how the parameters influence the precision of the tracker.

4.1 Measurement errors

The radar and camera measurements are noisy and show values that sometimes are far from the actual object's position. We first examine how the error $\|\mathbf{D^t} - \mathbf{D^t_{gt}}\|$ of the radar and camera measurements differ from the lidar (ground truth, gt). The errors are normalised such that the maximum error in the longitudinal axis of the radar is 1. As is shown in figure 4.1, the error of the camera is mainly in the longitudinal axis while the error of the radar is more uniformly spread out.



(A) Scatter plot of the error for radar measurements

(B) Scatter plot of the error for camera measurements

FIGURE 4.1: Scatter plots of sensor measurements errors

The standard deviation of the normalised errors is calculated (figure 4.2). As hypothesised, the radar has a smaller standard deviation than the camera in the longitudinal axis and vice versa in the latitudinal axis.

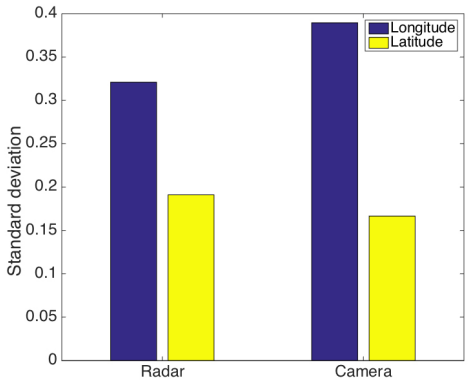


FIGURE 4.2: The standard deviation of (normalised) camera and radar measurements.

4.2 Parameter study

We will now investigate how the results of the method depend on the parameters β, γ, δ and ϵ. We do this by varying one parameter while keeping the rest fixed. Five different data sets were independently tried, giving a

standard deviation. The scores were normalised such that we get a score of one when the dynamic term is zero, $\beta = 0$.

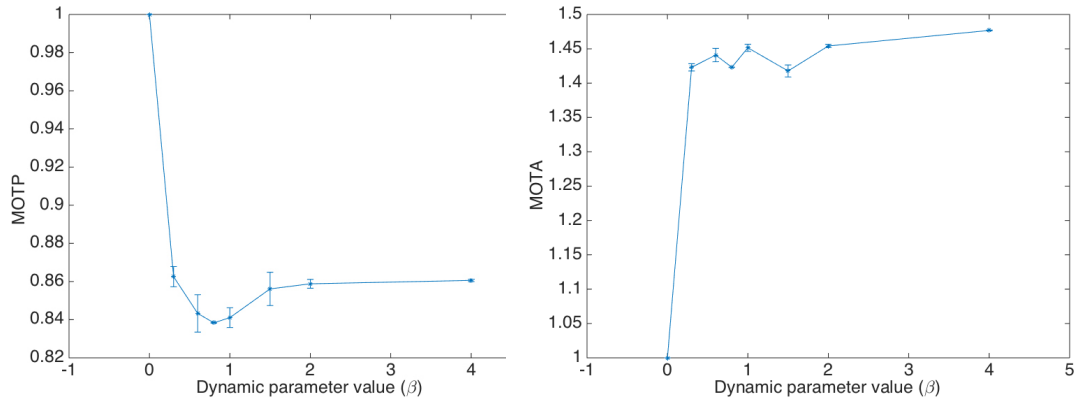


FIGURE 4.3: The MOTP (lower is better) and MOTA (higher is better) scores as functions of the dynamic parameter (β)

The dynamic parameter, β, is investigated (figure 4.3) by setting γ, δ and ϵ to zero while varying the value of β. A higher value for the dynamic parameter leads to the trajectories straightening out and the distance between subsequent points becoming more even. The MOTP score starts off at a high level when $\beta = 0$, then drops down and gradually gets worse when β is greater than 1. MOTA gets better with β greater than zero, and is then stable. It can be seen that an optimal value for β is around or slightly below 1. For the remaining results, $\beta = 1$, $\delta = 0$ and $\epsilon = 0$ is used.

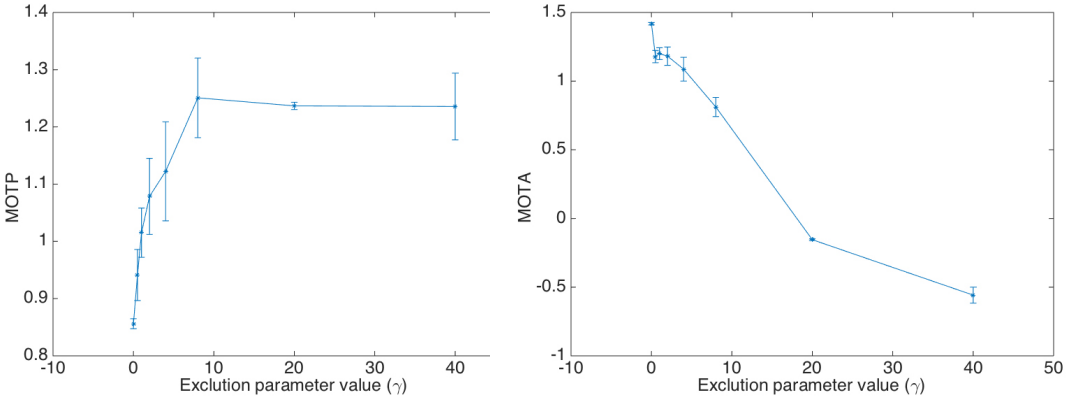


FIGURE 4.4: The MOTP (lower is better) and MOTA (higher is better) scores as functions of the exclusion parameter (γ)

The exclusion parameter, γ, is investigated by keeping β, δ and ϵ fixed and varying the value of γ (figure 4.4). The exclusion component of the energy keeps targets from occupying the same physical space by penalising points that are too close together. The MOTP score is the lowest with $\gamma = 0$ and increasing with higher γ. The MOTA score gets worse with higher values of γ. It can be seen that an optimal value for γ is 0. For the remaining results, $\gamma = 0$ is used.

The parameters for the persistence and regulation term did not change the score (not shown).

4.3 MOTP - Energy correlation

In order to investigate if there is a correlation between the MOTP score and the energy of the energy minimisation method, we plot the two values against the number of iterations of the gradient descent (figure 4.5). We see that there is a clear correlation between the two values. The MOTP score is normalised such that the graph starts at 1.

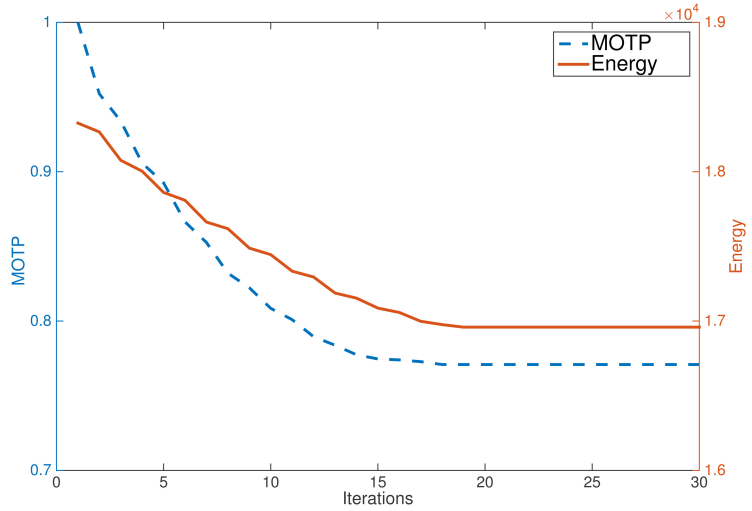


FIGURE 4.5: The energy and the MOTP score (normalised). Lower is better.

4.4 Method evaluation

The MOTP score is calculated for the individual sensors and for the output of the used energy minimisation method (figure 4.6). The values are normalised such that the energy minimisation gets a score of one. The energy minimisation is slightly improved over the mid-range radar which in turn has lower MOTP score than the camera and long-range radar. The energy minimisation results are better than the real-time method that is used currently (not shown in the graph).

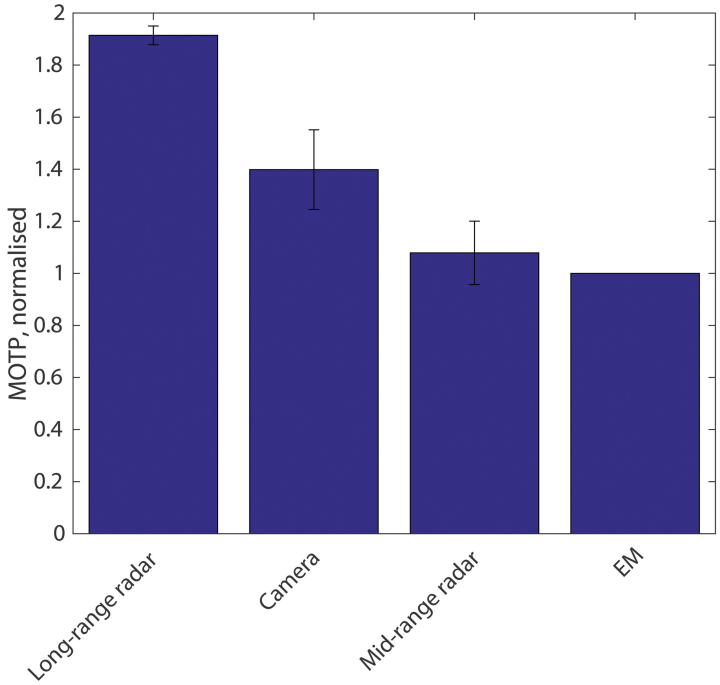


FIGURE 4.6: The MOTP score (normalised) for the used energy minimisation method and the individual sensors. Lower is better.

The MOTA score for the sensors and the result of the energy minimisation are compared (figure 4.7). The values are normalised to give the energy minimisation a score of one. All values are within one standard deviation of each other (including the currently used real-time method) so it is hard to draw any conclusions.

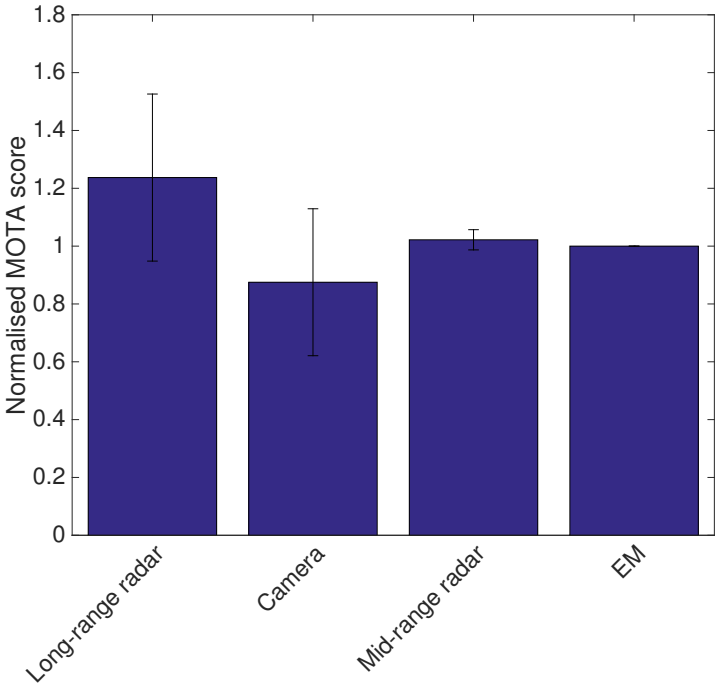


FIGURE 4.7: The MOTA score (normalised) for the used energy minimisation method and the individual sensors. Higher is better.

The false positive rate and miss rate are calculated (figure 4.8 and 4.9). The camera performs better in terms of false positive rate and worse in miss rate. The opposite is true for the long-range radar. The mid-range radar, energy minimisation and the currently used real-time method are within one standard deviation of each other.

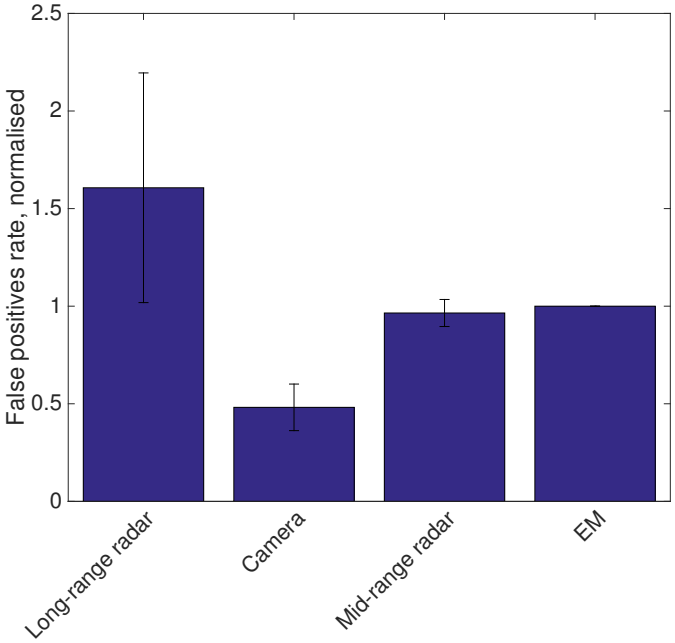


FIGURE 4.8: The false positive rate (normalised) for the used energy minimisation method and the individual sensors. Lower is better.

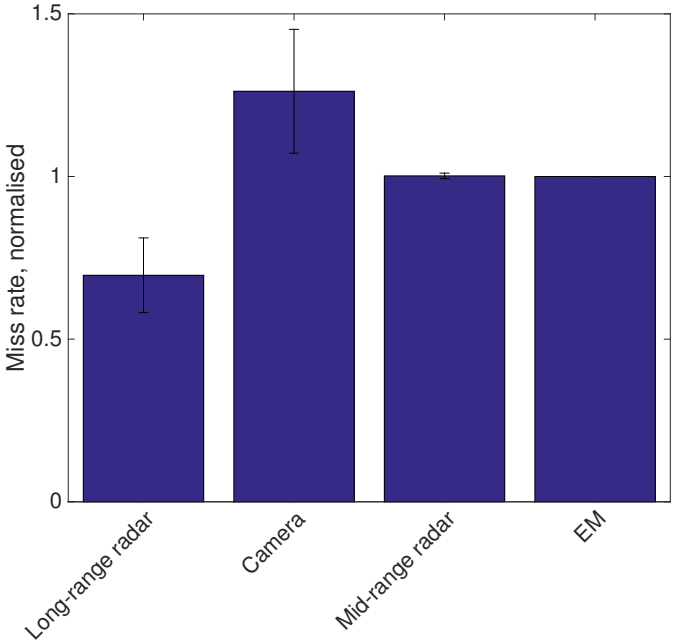


FIGURE 4.9: The miss rate (normalised) for the used energy minimisation method and the individual sensors. Lower is better.

Chapter 5

Discussion

In the discussion, conclusions are made from the results. Potential drawbacks and possible improvements are also discussed.

We have applied the energy minimisation method to sensor data collected with a camera and radar mounted in a car driving in a suburban traffic environment. Readings from the sensors were analysed which showed that the sensors have different characteristics. The results of this sensor analysis were used in the energy minimisation method. The data used was also preprocessed to better fit the method. Then the results were analysed using the MOTA and MOTP performance metrics which measure different aspects of the results.

5.1 Measurement errors

Based solely on the physical way that the camera and radar functions, the camera should have less errors in the azimuthal readings and the radar should be better at approximating the distance to the target. We examine if this is the case by comparing the readings of the lidar (used as ground truth) and the sensors, $\|\mathbf{D^t} - \mathbf{D^t_{gt}}\|$.

By doing this analysis we showed that the hypothesis of the characteristics was correct. The radar has a smaller standard deviation than the camera in the longitudinal axis and a smaller standard deviation in the latitudinal axis. This result was later used in the detection component of the energy. The longitudinal error of the radar got higher weights and the latitudinal errors got smaller weights. For the camera, the opposite was done; lower weights for longitudinal errors and higher weights for latitudinal errors.

5.2 Parameter study

The parameters were studied by keeping all but one fixed, and varying the value of the last parameter. The dynamic term was the first investigated. This term considers the relative velocities between all points. The difference in velocity between subsequent points should be small in order for the dynamic term to be low. When increasing the dynamic term parameter, β, the trajectories tend to be straightened out and the distance between subsequent points become more equal.

We showed that increasing the dynamic parameter improves the MOTP results up to a certain point, after which the results become worse. The MOTA score also improves for values above zero. This means that we improve trajectories by having a smoothing term in the energy. The dynamic term both limits the movements of other vehicles and filters out some noise in the data.

The exclusion term limits how small distance is accepted between two different targets; this prevents two vehicles from occupying the same physical space. The exclusion component of the energy goes to infinity as the distance between two targets goes to zero, and quickly falls off when the distance between two targets is physically possible.

A higher value than zero for the exclusion parameter was shown to worsen our results, both in the MOTP and MOTA scores. A possible reason for this could be that, unlike the scenario that Milan et. al studied, we rarely have targets being very close to one another. This limits the possible benefits of the exclusion component. In rare cases, we can have one object being falsely interpreted as two objects. In these cases, the exclusion term will strongly push these two apart, making them both go far away from the ground truth.

Both the persistence and the regulation components were shown not to influence the results of the method. The persistence term works by moving the first and the last point of each trajectory closer to the edge of the field of view. The reason for this is that other vehicles cannot appear suddenly in the middle of the field of view - they must come from somewhere. The regulation term simply tries to limit the total number of tracked objects and the number of points in the trajectories. Both of these components have little to no effect on the end result. One reason for this could be that

the regulation term only changes value when points are added or removed (and not when points are moved). The persistence component affects a very small proportion of all points. It also requires a clearly defined border of the field of view while the radar's field of view border is more diffuse.

5.3 MOTP - Energy correlation

One of the main questions about the energy minimisation approach is wether or not the calculated energy is corresponding to how close the trajectories are to reality. We investigate this by comparing how the energy and MOTP decrease during the gradient descent. The MOTP metric shows how well the calculated trajectory corresponds to the ground truth by giving a lower score to trajectories closer to the ground truth (lidar). We see that there is a clear correspondence between the energy and the MOTP score (figure 4.5).

5.4 Method evaluation

In order to see how well the used energy minimisation method performs, we compare it to the individual sensors as well as with the currently used real time tracker. The computed trajectory is compared to the ground truth (lidar), both with the MOTA and the MOTP performance metrics.

We showed that our method performs better than the currently used method and is better or on par with the individual sensors in all aspects. The MOTP score is on par with the mid-range radar and is better than the camera, long-range radar and the currently used real time tracker. The MOTA metric is a combination of false positive rate, miss rate and the mismatch error rate. The MOTA score of our method is on par with the individual sensors as well as the currently used real time method.

The false positive rate measures how often non existing targets are perceived to exist and the miss rate is how often existing targets are not found. The long-range radar performs better than the other sensors in terms of false positive rate and worse in miss rate. The opposite is true for the camera. It is not surprising that sensors that are good at false positive rate are worse with miss rate since these two scores are naturally connected. If a

sensor is overly sensitive, then it will not miss many targets but it will on the other hand make many false positives. Our method and the mid-range radar are scoring somewhere in between the camera and the long-range radar.

5.5 Comments

In the scenario considered by Milan et al., a camera is pointed downwards towards walking pedestrians, which results in a clearly defined field of view. When using radars the field of view is more diffuse towards the edges, which makes the persistence component of the energy less important.

Milan et al. were able to use an appearance term in the energy function. This appearance energy was calculated for all positions in the field of view. The camera data used in our implementation consisted of only positions of other objects since calculating the appearance energy for each position was not possible. The radar provides a measure of the radar cross-section of all detected objects. In order for us to use this data we would need to make an association between detections and trajectories, which would defeat one of the advantages of the method.

The data used is gathered in a suburban traffic environment. Other types of environments may influence the typical movements of other vehicles and the parameters may therefore need to be tweaked for these.

The pre-processing performed on the sensor measurements, described in section 3, affects the results. Particularly the misses component of the MOTA-score is influenced by the pre-processing. The sensor data that is used has also been processed in the sensor (likely with a Kalman filter). While the errors of raw sensor measurements often are independent, the errors in the data we use are not, which may have influenced our results.

The jump moves described in section 3.2 were implemented but deemed too computationally time consuming to be used in the final results.

5.6 Future work

The energy function that is used in the energy minimisation method we consider is composed of several different terms. The method can thus easily be modified and improved by adding, removing or tweaking these components. For example, a term that prevents our moving objects from overlapping with static obstacles could be included.

The dynamic term that we use restrains the objects' acceleration in all directions equally. This dynamic model is likely good for pedestrians which can move quite erratically. A better dynamic model for vehicles would consider the change in yaw and change in speed. A constant turn rate model is a popular choice [24].

The exclusion model used considers the shape of the vehicles to be circles. A more realistic shape could be used; this would however add complexity since the orientation of each vehicle would need to be considered.

5.7 Conclusions

We have seen that the energy minimisation method proposed by Milan et al. is able to improve some aspects in comparison to the individual sensors and the currently used real-time tracker. The energy function used in the method clearly corresponds to how close the trajectory is to reality. The improvements are seen in the MOTP score while the MOTA score and its components are similar to the mid-range radar.

Bibliography

[1] Anton Andriyenko and Konrad Schindler. "Multi-target tracking by continuous energy minimization". In: *Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on*. IEEE. 2011, pp. 1265–1272.

[2] Klaus Bengler et al. "Three decades of driver assistance systems: Review and future perspectives". In: *IEEE Intelligent Transportation Systems Magazine* 6.4 (2014), pp. 6–22.

[3] Keni Bernardin and Rainer Stiefelhagen. "Evaluating multiple object tracking performance: the CLEAR MOT metrics". In: *EURASIP Journal on Image and Video Processing* 2008.1 (2008), pp. 1–10.

[4] Duc Phu Chau, François Bremond, and Monique Thonnat. "A multi-feature tracking algorithm enabling adaptation to context variations". In: *Imaging for Crime Detection and Prevention 2011 (ICDP 2011), 4th International Conference on*. IET. 2011, pp. 1–6.

[5] Lars Danielsson. "Tracking and radar sensor modelling for automotive safety systems". PhD thesis. Chalmers University of Technology, 2010.

[6] Q. Gan and C.J. Harris. "Comparison of two measurement fusion methods for Kalman-filter-based multisensor data fusion". In: *IEEE Transactions on Aerospace and Electronic Systems* 37.1 (Jan. 2001), pp. 273–279. URL: http://ieeexplore.ieee.org/document/913685/.

[7] Neil J Gordon, David J Salmond, and Adrian FM Smith. "Novel approach to nonlinear/non-Gaussian Bayesian state estimation". In: *IEE Proceedings F-Radar and Signal Processing*. Vol. 140. 2. IET. 1993, pp. 107–113.

[8] Rudolph Emil Kalman. "A new approach to linear filtering and prediction problems". In: *Journal of basic Engineering* 82.1 (1960), pp. 35–45.

[9] Pavlina Konstantinova, Alexander Udvarev, and Tzvetan Semerdjiev. "A study of a target tracking algorithm using global nearest neighbor approach". In: *Proceedings of the International Conference on*

Computer Systems and Technologies (CompSysTech'03). 2003, pp. 290–295.

[10] Delphi Automotive LLP. *Delphi Electronically Scanning Radar*. http://www.delphi.com/manufacturers/auto/safety/active/electronically-scanning-radar. Accessed: 2016-10-09.

[11] Christian Lundquist. "Sensor fusion for automotive applications". PhD thesis. Linköping University Electronic Press, 2011.

[12] Volvo Cars Owner Manual. *Collision warning - Pedestrian detection*. http://support.volvocars.com/en-CA/cars/Pages/owners-manual.aspx?mc=y286&my=2016&sw=15w17&article=d3d274ecedf1c586c0a801e8004927e7. Accessed: 2016-10-09.

[13] Hermann Winner Michael Darms. "A Modular System Architecture for Sensor Data Processing of ADAS Applications". Proceedings of the 2008 Intelligent Vehicles Symposium, Las Vegas, NV (pp. 729–734). In: *IEEE* (2008).

[14] Anton Milan. "Energy Minimization for Multiple Object Tracking". PhD thesis. TU Darmstadt, 2014.

[15] Anton Milan, Stefan Roth, and Konrad Schindler. "Continuous energy minimization for multitarget tracking". In: *IEEE transactions on pattern analysis and machine intelligence* 36.1 (2014), pp. 58–72.

[16] Thi Lan Anh Nguyen, Duc Phu Chau, and Francois Bremond. "Robust global tracker based on an online estimation of tracklet descriptor reliability". In: *Advanced Video and Signal Based Surveillance (AVSS), 2015 12th IEEE International Conference on*. IEEE. 2015, pp. 1–6.

[17] Prachi Patel. *Will We Prove That Autonomous Cars Are Safe Before They Go on Sale?* http://spectrum.ieee.org/cars-that-think/transportation/self-driving/we-might-never-have-proof-that-autonomous-cars-are-safe. Accessed: 2016-10-09.

[18] RH Rasshofer, M Spies, and H Spies. "Influences of weather phenomena on automotive laser radar systems". In: *Advances in Radio Science* 9.B. 2 (2011), pp. 49–60.

[19] Mikel Rodriguez et al. "Density-aware person detection and tracking in crowds". In: *Computer Vision (ICCV), 2011 IEEE International Conference on*. IEEE. 2011, pp. 2423–2430.

[20] Philip E. Ross. *Helsinki Tries Self-Driving Buses in Real Traffic*. http://spectrum.ieee.org/cars-that-think/transportation/self-driving/helsinki-tries-selfdriving-buses-in-real-traffic. Accessed: 2016-10-09.

[21] Philip E. Ross. *Tesla's Massive New Autopilot Update Is Released, Promising Safer Driving*. http://spectrum.ieee.org/cars-that-think/transportation/self-driving/teslas-massive-new-autopilot-update-is-released. Accessed: 2016-10-09.

[22] Julian Ryde and Nick Hillier. "Performance of laser and radar ranging devices in adverse environmental conditions". In: *Journal of Field Robotics* 26.9 (2009), pp. 712–727.

[23] Xu Yan, Ioannis A Kakadiaris, and Shishir K Shah. "What do I see? Modeling human visual perception for multi-person tracking". In: *European Conference on Computer Vision*. Springer. 2014, pp. 314–329.

[24] Guan Zhai, Huadong Meng, and Xiqin Wang. "A constant speed changing rate and constant turn rate model for maneuvering target tracking". In: *Sensors* 14.3 (2014), pp. 5239–5253.