T.C.

ALTINBAS UNIVERSITY

Institute of Graduate Studies

Electrical and Computer Engineering

OBJECT SEGMENTATION AND TRACKING IN DIGITAL IMAGE WITH ACTIVE CONTOUR MODEL AND HEURISTIC METHOD

Omar Mohammed AL-QARAGHULI

Master of Science

Supervisor

Asst. Prof. Dr. Abdullahi Abdu Ibrahim

Istanbul, 2020

OBJECT SEGMENTATION AND TRACKING IN DIGITAL IMAGE WITH ACTIVE CONTOUR MODEL AND HEURISTIC METHOD

By

Omar Mohammed AL-QARAGHULI

Electrical and Computer Engineering

Submitted to the Institute of Graduate Studies

in partial fulfillment of the requirements for the degree of

Master of Science

ALTINBAŞ UNIVERSITY

2020

The thesis titled **“OBJECT SEGMENTATION AND TRACKING IN DIGITAL IMAGE WITH ACTIVE CONTOUR MODEL AND HEURISTIC METHOD”** prepared and presented by **Omar Mohammed AL-QARAGHULI** was accepted as a Master of Science Thesis in Electrical and Computer Engineering.

____
Asst. Prof. Dr. Abdullahi Abdu Ibrahim
Supervisor

Thesis Defense Jury Members:

Asst. Prof. Dr.	School of Engineering and Natural Sciences, Altinbas University	____________
Prof. Dr.	School of Engineering and Natural Sciences, Altinbas University	____________
Prof. Dr.	Faculty of Engineering and Architecture, Nisantasi University	____________

I certify that this thesis satisfies all the requirements as a thesis for the degree of Master of Science.

Approval Date of Institute of Graduate Studies: ____/____/____

I hereby declare that all information in this document has been obtained and presented in accordance with academic rules and ethical conduct. I also declare that, as required by these rules and conduct, I have fully cited and referenced all material and results that are not original to this work.

Omar Mohammed AL-QARAGHULI

DEDICATION

I devote and pledge this research work to my supervisor who is salient for guiding me through whole research work as well as my family for always assisting me in my hard time.

ACKNOWLEDGEMENTS

First and foremost, I would like to thank my supervisor Asst. Prof. Dr. Abdullahi Abdu Ibrahimfor guiding and helping me along the way in writing this dissertation. Discussing my progress, problems, and ideas with my supervisor Asst. Prof. Dr. Abdullahi Abdu Ibrahima couple of times every week helped me tremendously in understanding the logic behind the research. It made me better realize the technical need for this research work. He helped me by looking outside of all the machine learning and network-based information and considering the cryptocurrency-based knowledge from a more holistic perspective than just from the productivity point of view. He patiently explained to me a great number of technical details with regards to block-chain and cryptocurrency, for example how to deal with solving problems and to provide insight into the reliability of those results. Without his time and help none of this research would have been possible, and I'm grateful to him for giving me the opportunity to work on this interesting subject. I am also very thankful to my parents as well as my friends who always supported me through whole of this journey.

ABSTRACT

OBJECT SEGMENTATION AND TRACKING IN DIGITAL IMAGE WITH ACTIVE CONTOUR MODEL AND HEURISTIC METHOD

AL-QARAGHULI, Omar Mohammed,

M.Sc., Electrical and Computer Engineering, Altınbaş University,

Supervisor: Asst. Prof. Dr. Abdullahi Abdu Ibrahim

Date: Aug/2020

Pages: 70

In this thesis, object segmentation and tracking in digital image with Active Contour model and heuristic method and along with it multiple ground target tracking algorithms are studied. From different aspects of the ground target tracking, three different types of tracking algorithms are proposed according to the specialties of the ground target motion and sensors employed. Firstly, the dependent target tracking for ground targets is studied. State dependency is a common assumption in traditional target tracking algorithms, while this may not be the true in ground target tracking as the motion of targets are constraint to certain path. This approach largely reduces the exhaustive searching in common state-of-art trackers while maintains efficient representation of the target appearance change. The primary contribution of this project will be a method to propose a simple approach to take advantage of state-of-the-art methods to apply to video object segmentation problem using active contours with heuristic methods for reappearance detection. Experiments on 100 public benchmark videos, as well as a high frame rate benchmark, are carried out to compare the performance with the state-of-art published algorithms. The results of the experiment show the proposed tracker achieves good performance while beats other algorithms in speed with a large margin. The proposed visual target tracker is integrated into a new multiple ground target tracking algorithm using a single camera. The multi-target tracker addresses the issues in the target detection, data association and track management aside from the single target tracker. A perspective aware detection algorithm utilizing the recent advanced Convolutional Neural Networks (CNN) based detector is proposed to detect multiple ground targets and alleviate the weakness of CNN detectors in detecting small objects. A hierarchical class tree based multi-class data association is presented to solve the multi-class association problem with potential misclassified detections. Track management is also improved utilizing the high efficiency detectors and a Support Vector Machine (SVM) based track deletion is proposed to correctly remove the dead tracks. Benchmarking is presented in experiments and results are analyzed. A case study of applying the proposed algorithm is provided demonstrating the usefulness in real applications.

Keywords: Convolutional Neural Networks, Object Tracking, Support Vector Machine (SVM), tracking algorithm.

TABLE OF CONTENTS

Pages

LIST OF TABLES

Pages

LIST OF FIGURES

Pages

LIST OF ABBREVIATIONS

AI	:	Artificial Intelligence
BIS	:	Bank of International Settlement
CPP	:	Cryptocurrency Price Prediction
DLT	:	Distributed Ledger Technology
FTR	:	Funds Transfer Regulation
FIR	:	Fixed Investment Rate
LCL	:	Legal Cryptocurrency Law
ML	:	Machine Learning
OTF	:	Organized Trading Facility
PoS	:	Proof of Stake
PoW	:	Proof of Work
TCI	:	Technical Trade Indicator

1. INTRODUCTION

The Modern world has been becoming more and more advance in terms of technologies and these technologies are making our lives more comfortable and easier. The object segmentation and tracking in digital image can be done by using active contour model which is a framework of computer vision and it can be implemented by heuristic method which is a problem solving method. In this thesis, the various developments in technology over the last two decades have greatly increased our ability to conduct experiments and obtain data at the cellular and molecular levels [1]. This growth can be attributed to the development of a range of fluorescent probes and the milestones that have been achieved in the field of optical microscopy. The improvements in optical microscopy provide us the insight that is necessary to understand biological processes at the molecular and cellular level. The availability of a wide range of fluorescent probes provides us more control over the visualizations of experiments and helps us focus on particles of interest [3]. These advancements provide us access to a plethora of experimental data; however being able to separate data of interest from the entire set is still a challenge. Object tracking plays a major role in studying the movement and behavior of proteins along the cell membrane, aids us in studying the nature of the motion of particles at the cellular level. Knowledge of the nature of motion of particles equips us to better understand biological processes and cellular kinetics. The particles of interest are usually labeled with the help of an optical label or a fluorescent tag, which helps distinguish these from others that may also be present in the medium. Once these particles are easily identifiable the next step in the process is to track them. Tracking a particle involves observing the motion of particles in the medium by estimating the positions of the particles during this time. Tracking of particles is typically a very laborious process since it is complicated by the fact that there are number of factors which are not under our control. The diversity in the type, size and shapes of particles found in the abundant biological processes makes it hard to have a uniform detection scheme. In case of 2D imaging, particles tend to move in and out of the plane of focus which leads to temporary disappearance and reappearance of particles depending on the nature of the process itself particles may merge or split; typically the density of the particles of interest is also not under the control of the experimental set-up of different approaches of object tracking. The above mentioned constraints make it hard to have a universal solution that detects and tracks particles for all, or even most of the applications under our interest. In order to accurately detect and track particles, algorithms will have to be developed for the specific application that is being investigated. This thesis focusses on developing and validating an algorithm for tracking motor proteins moving along the axons. A detection algorithm was developed and Linear Assignment Problem (LAP) [5] based tracking algorithm was used to track motor proteins. A comparative study of some of the existing algorithms was also performed as a part of this study [2].

The various developments in technology over the last two decades have greatly increased our ability to conduct experiments and obtain data at the cellular and molecular levels. This growth can be attributed to the development of a range of fluorescent probes and the milestones that have been achieved in the field of optical microscopy. The improvements in optical microscopy provide us the insight that is necessary to understand biological processes at the molecular and cellular level. The availability of a wide range of fluorescent probes provides us more control over the visualizations of experiments and helps us focus on particles of interest. These advancements provide us access to a plethora of experimental data; however, being able to separate data of interest from the entire set is still a challenge. Particle tracking plays a major role in studying the movement and behavior of proteins along the cell membrane, aids us in studying the nature of the motion of particles at the cellular level. Knowledge of the nature of motion of particles equips us to better understand biological processes and cellular kinetics [9]. Motor proteins are responsible for the transport of cellular cargo and malfunctions in their behavior are linked to multiple human diseases [6].

The particles of interest are usually labeled with the help of an optical label or a fluorescent tag [8], which helps distinguish these from others that may also be present in the medium. Once these particles are easily identifiable the next step in the process is to track them. Tracking a particle involves observing the motion of particles in the medium by estimating the positions of the particles during this time. Tracking of particles is typically a very laborious process since it is complicated by the fact that there are number of factors which are not under our control. The diversity in the type, size and shapes of particles found in the abundant biological processes makes it hard to have a uniform detection scheme. In case of 2D imaging, particles tend to move in and out of the plane of focus which leads to temporary disappearance and reappearance of particles [7] depending on the nature of the process itself particles may merge or split; typically, the density of the particles of interest is also not under the control of the experimental set-up. The above mentioned constraints make it hard to have a universal solution that detects and tracks particles for all, or even most of the applications under our interest. In order to accurately detect and track particles, algorithms will have to be developed for the specific application that is being investigated. This thesis focusses on developing and validating an algorithm for object segmentation and tracking in digital image.

A detection algorithm was developed and Linear Assignment Problem (LAP) [1] based tracking algorithm was used to track motor proteins. A comparative study of some of the existing algorithms was also performed as a part of this study [2].

1.1 ACTIVE CONTOUR MODEL USAGE IN OBJECT TRACKING:

Active Contour model is a computer vision frame work which in this thesis we are using for object segmentation and Object Tracking in Digital image. The increasing need for object segmentation, and Object Tracking there is still a shortage for algorithms and methods to resolve it. Most of current segmentation methods, however, target a single image, a single frame instead of sequences of frames Therefore, there are some activities created to encourage the development of more efficient algorithms, methods to do object segmentation task That's why Active Contour model is one of the best frame work of computer vision that can be used in Object Segmentation and object Tracking in Digital Image. Object segmentation is about finding objects and creating a boundary around each object as tightened as possible. With this ability, its applications include tracking objects of interest given limited information. During the authors' study of the problem of object segmentation, the authors come across object saliency papers [1] [2] and find that these papers can lead to a fresh approach, the one that is not applied widely, to object segmentation task.

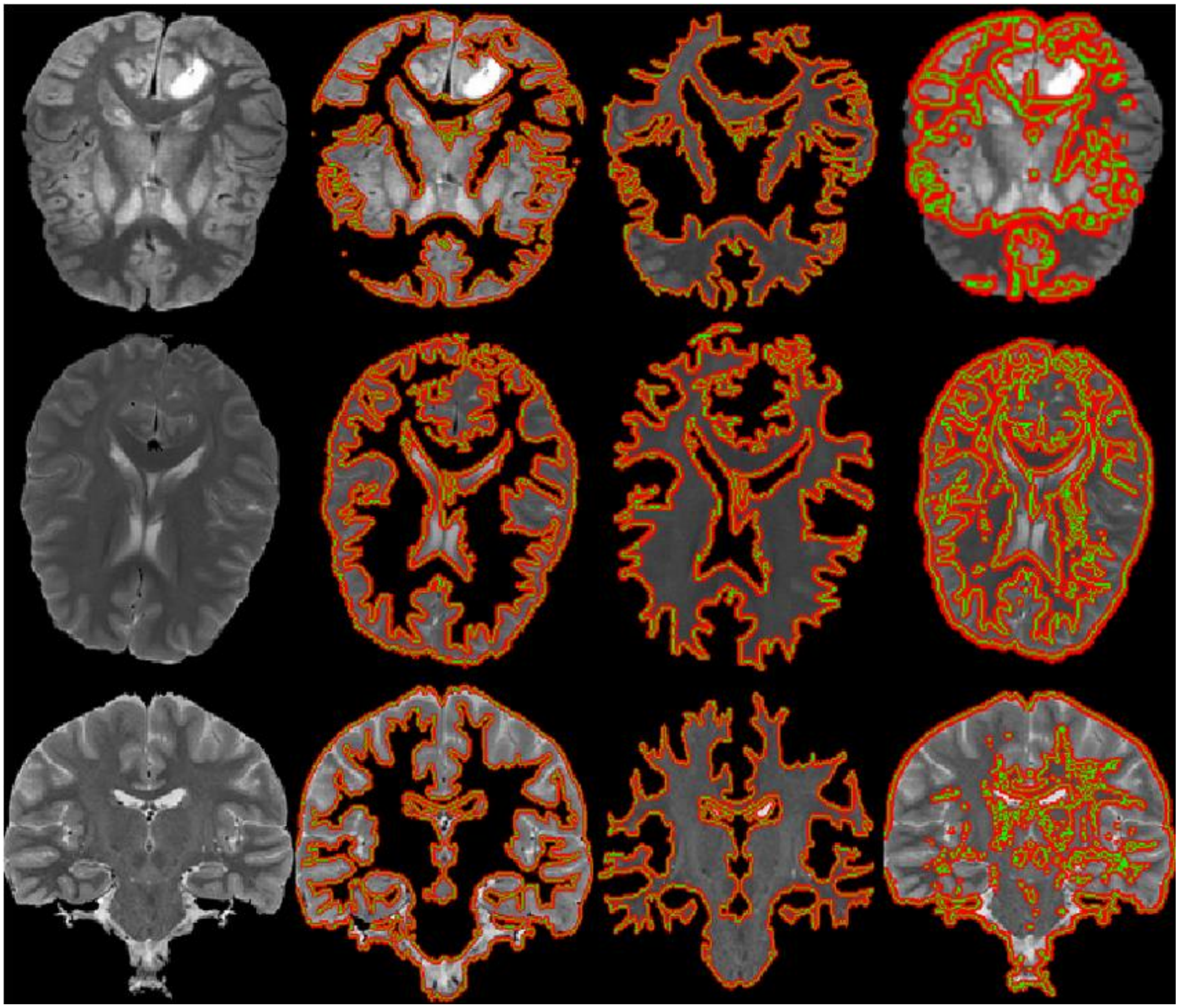


Figure 1.1: Segmentation results of the Active Contour Model [2]

The active contours of an object are the state by which that objects is easily noticeable in comparison with its neighbors. A good object active contour method can not only give the information about how much the object stands out but also extracts fairly accurately the object from its background [2] with heuristic approach and training through machine learning. Currently, once the Digital images have been obtained, kymograph of the series of images is generated.

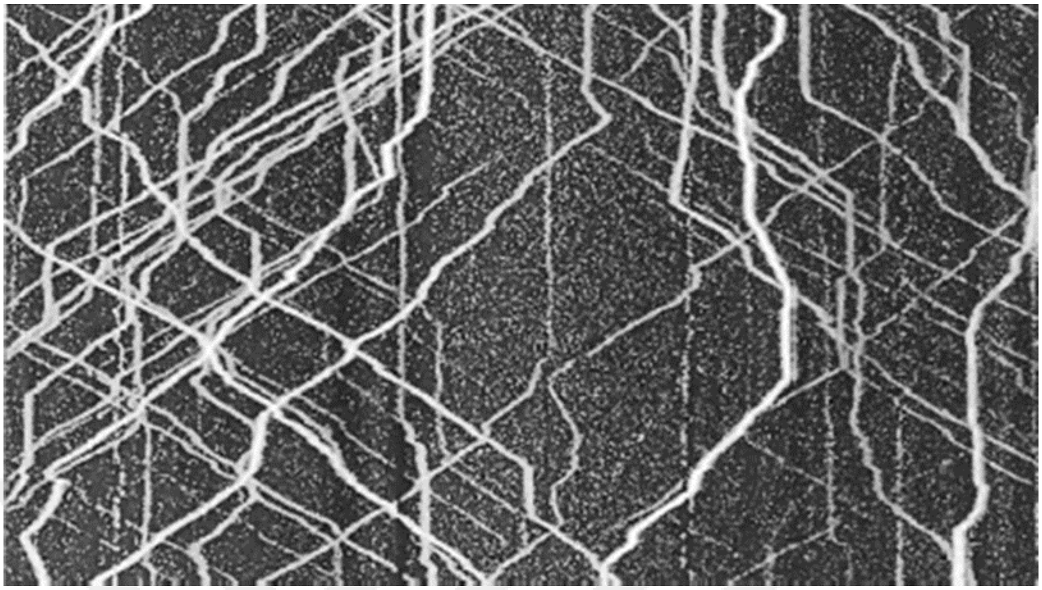


Figure 1.2: Kymograph [3]

A kymograph is a graph which contains the estimates of the position of the particles versus time [3]. Once the kymograph is generated, the required statistics are calculated manually by observing the kymograph. This is a laborious and time consuming process. Also, there is no way to validate the results as the results rely on observation and are likely to be subjective. This is the motivation behind working on this topic as part of the dissertation.

1.2 OBJECT TRACKING WITH HEURISTIC APPROACH

The heuristic Approach is to ease the difficulties associated with the manual tracking process Because it is a problem solving approach that give us the solutions and allow us to come up with a suitable algorithm to successfully track the particles and obtain results that are as accurate as or more accurate than the results obtained by manual tracking with active contour model in heuristic approach first we identify the problem then we come up with a hypothesis in which we gather data and later analyze it . In this project to automate the detection and tracking of particles and generate the required statistics as was being generated by the manual tracking. The relevant results that had to be generated were same as the ones generated from manual tracking. The other main benefit of this approach in this thesis is to propose a simple approach to take advantage of state-of-the-art methods to apply to object segmentation problem using active contour model and heuristic method for reappearance detection. The output of the approach should be a ready-to-use matte that can be applied to compose two scenes serving automatic segmentation purpose. Ultimately, the we need to create a demo application for demonstrating the ability to apply object segmentation to video matting. For the implementation of this work, python programming will be used for training and testing purpose with pre-trained models.

1.3 AIMS OF THESIS

The aim of the Thesis is to track an object, in Digital images using active contour model and heuristic method that are explained above, the digital image will acquire by a Digital camera, and to simultaneously estimate its pose, knowing a 3D model of it. However, it is difficult to obtain high-resolution images in digital camera due to the other limitations. Single image super-resolution (SISR) method can generate a high-resolution (HR) image from a single low-resolution (LR) image. In this project, we aim to address mainly the following issues:

- Deep learning methods, more specifically object tracking imaging with digital camera using these two approaches.
- Finding the optimal choice of CNN architecture and hyper parameters of the training process.

1.4 CONTRIBUTIONS

The primary contribution of this project will be a method to propose a simple approach to take advantage of state-of-the-art methods to apply to video object segmentation problem using active contours with heuristic methods for reappearance detection. The output of the approach should be a ready-to-use matte that can be applied to compose two scenes serving automatic matting purpose. Ultimately, the authors create a demo application for demonstrating the ability to apply object segmentation to video matting.

Approaches to follow:

- Machine learning for training
- Active contours as model
- Heuristic methods for reappearance detection

- Digital image processing

1.5 PLANNING OF CHAPTERS

The planning of chapters in this thesis will be like this ***Chapter 1: Introduction*** This section presents an overview of the thesis, as well as a short summary off each chapter. ***Chapter 2: Theoretical Background*** This chapter presents background knowledge and related work that is relevant to the topics of the thesis. The object segmentation and object tracking will be done in this thesis with active contour model and heuristic method, as well as different ways to track the objects. Following this, several topics related to object tracking will be describe

Chapter 3: Methodology This chapter presents a theoretic overview of the method proposed in this thesis. It presents the structure of the method and the proposed framework for implementation and comparison, and suggests several sub-methods that can be used to implement it. A dataset of this object tracking in digital image will be used to train in Python and evaluate the proposed method. This dataset will also be presented.

Chapter 4: Results The results and lab practice are presented and compared in this chapter.

Chapter 5: Discussion and Conclusion The final chapter presents a summary and conclusion of the thesis. This chapter provides a discussion off the results, as well as the considerations and implications of the thesis, and a description of future research that may be conducted to further study the subjects presented, and in last there will be References.

2. LITERATURE REVIEW

Object tracking plays a major role in studying the movement and behavior of proteins along the cell membrane, aids us in studying the nature of the motion of particles at the cellular level. Knowledge of the nature of motion of particles equips us to better understand biological processes and cellular kinetics. The particles of interest are usually labeled with the help of an optical label or a fluorescent tag, which helps distinguish these from others that may also be present in the medium. Once these particles are easily identifiable the next step in the process is to track them. Tracking a particle involves observing the motion of particles in the medium by estimating the positions of the particles during this time. Tracking of particles is typically a very laborious process since it is complicated by the fact that there are number of factors which are not under our control. The diversity in the type, size and shapes of particles found in the abundant biological processes makes it hard to have a uniform detection scheme. In this chapter, we organize the literature survey in a few sections. In the first section, we review the object segmentation with heuristic methods and related concepts. In section two, we will discuss about object tracking. In section three, we outline the video-based objects along with image based objects and then specifically object tracking by contour model.

2.1 OBJECT SEGMENTATION

As has been stated by [33], US imaging can be subdivided into multiple modes. The A-mode thereby stands for a 1-dimensional line scan where a single transducer emits sound waves into the body, recoding the echoes. The B-mode is a 2-dimensional mode using a linear array of transducers that allow for a planar scan. The most important mode concerning this thesis is the M-mode, that utilizes a successive A or B-mode scan, creating an infinite discrete stream of US frames. According to [33], modern US systems use improved concepts such as the Doppler mode that allows for a visualization of blood flow. Sending the sound waves at different angles provides a basis for 3-dimensional US scans. However, due to the fact that only the M-mode in combination with B-mode imaging allows for an infinite 2D video stream, which is also the most commonly used mode in diagnostic ultrasonography, this thesis focuses only on the B-M mode [33].During the generation of the US images, the physics of the emitted sound waves allow for the emergence of artifacts in the produced images. According to [22], these artifacts comprise reverberation effects, ring down artifacts, mirrored objects, reflections, acoustic shadows, enhancement artifacts and statistical noise. All of those artifacts affect the US image quality and increase the difficulty of object segmentation as well as object tracking. Hence, they increase the amount of preprocessing that is needed and for this reason their presence may also increase the required processing time t-frame for a single frame. Regarding to [21], reverberation effects are caused by sound that is bouncing between tissue boundaries before returning to the receiver and they emerge as equally spaced lines along the ray direction. Resonating air bubbles can produce additional sound that irritates the system and leads to unusually bright areas at the bottom of these bubbles, called ring down artifacts. The enhancement artifact is similar to the ring down effect but caused by sound traveling through various mediums with different attenuation. Reflections and mirror images both create duplicated structures at misleading positions in the resulting images. Those artifacts may pose the most critical threat in medical terms and can only be detected by utilizing additional knowledge (e.g. kidney stones can only appear inside the kidney region). The attenuation or shadow artifact produces darker regions below strongly attenuating objects such as kidney stones. However, this effect may not only harmfully modify the image but can also be used for identifying such objects. Speckle noise is a random, deterministic and statistically distributed interference pattern in US images that affects the visible image quality and complicates object tracking through movement detection. Some of the potentially emerging artifacts are depicted in Figure 2.1

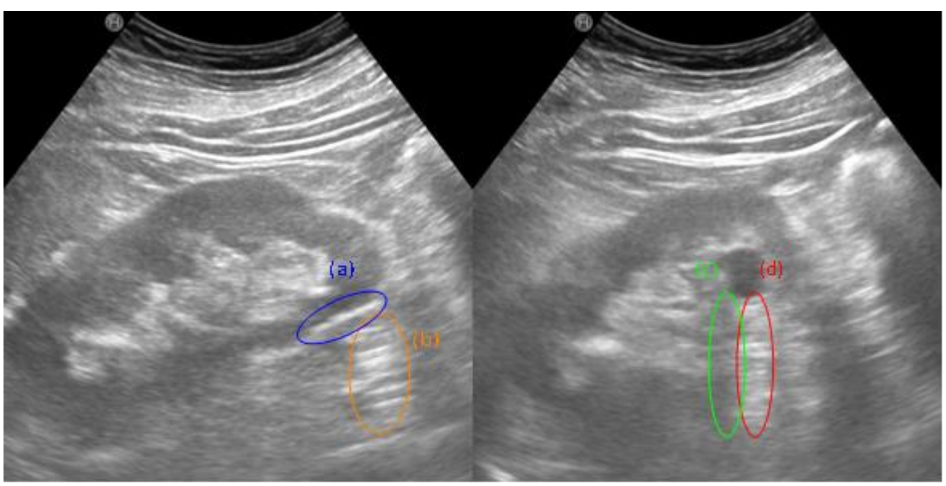


Figure 2.1: Artifacts in US images of the kidney region, showing ring down (a), reverberation (b), attenuation (c) and enhancement (d) effects.

2.1.1 Template Matching

The complexity of object tracking is influenced by many factors. As has been stated in Chapter 2 of this thesis, US video streams comprise a large number of artifacts that lower the image quality. Whereas video surveillance applications may deal with a static background and moving target objects, the US video application includes at least patient movement such as respiration which conducts a non-linear transform on the frames, thus leading to a separate movement of both, background and target objects. Additionally, the transducer may be moving as well, leading to background movement, partially visible objects or objects moving completely outside the capture region. Since a B-scan of the US modality is a planar scan, a movement of the transducer may position the sectional plane such that it does not intersect with the target object anymore whereas the background remains mostly constant. Many of the presented applications also involve creating a model of the desired object. Although, this can be a simple task for applications such as player tracking in sports casts using static templates, deformable objects in US videos require dynamic models, thus complicating the modeling task. Other factors that exacerbate the tracking process comprise occlusion effects when tracking multiple objects. Player objects in sports video streams may be subject to occlusion as described by [40]. However, these occlusion effects may also occur in US videos, depending on the application. Despite the differences between the listed applications, common characteristics can be extracted. All of them start with a set of input frames F-input in a finite or infinite video stream. According to the type of application and the involved artifacts, pre-processing may be required e.g. to remove noise, hence this step is optional. It is followed by a segmentation of objects Oseg = {o1, ..., on}, n ∈ N and a subsequent extraction of features such as edges or histograms according to the objects. These two stages may be repeated multiple times in order to refine the results. Object based image analysis (OBIA) as described by [19] could be seen as multiple iterations of segmentation and feature extraction. Afterward, the extracted object descriptions and related features are submitted to a model that is used to select relevant objects and predict their future locations. In a last step, the predicted objects are being marked in output frames F-output ⊆ F-input. A visualization of this abstract process scan be found in Figure 2.2. This pipeline is similar to the image engineering frame work

as has been stated [42], consisting of image processing, image analysis and image understanding. Pre-processing or image processing thereby represents the lowest layer, followed by a middle layer consisting of image segmentation, object detection and feature extraction algorithms which are part of image analysis. The highest level, named image understanding, eventually comprises the modeling and prediction tasks.

However, according to [25] this pipeline represents only one of two different classes of trackers, namely bottom-up approaches. There is also the class of top- down trackers which utilize a sort of reverse approach to the bottom-up trackers. Thus, a top-down approach generates object hypotheses from an initial model first and tries to verify them by using the input images F-input before marking the objects in the output images F-output and updating the model. A visualization of a top-down pipeline is shown in Figure 2.2.

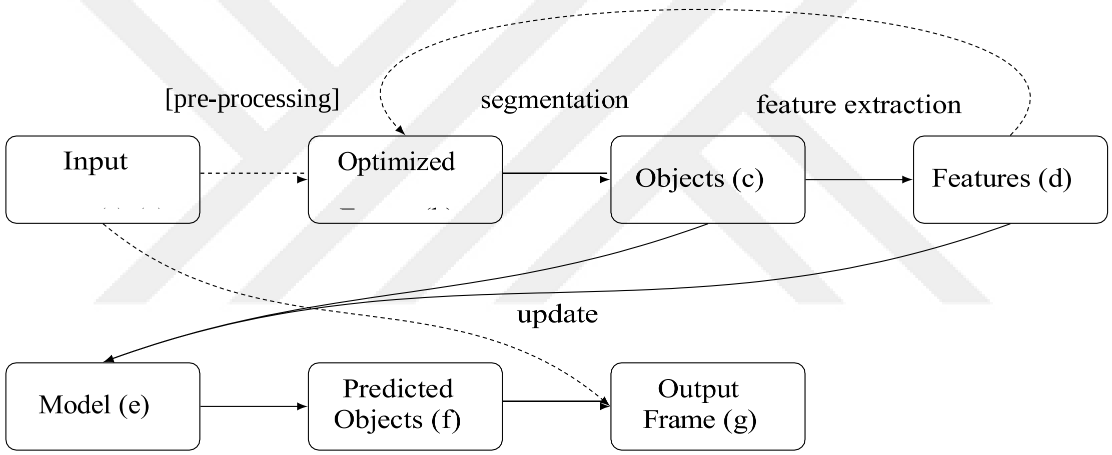


Figure 2.2: Generic bottom-up object tracking pipeline, showing the processing steps between the input frame (a) and the output frame (g) including optional stages such as pre-processing (a b).

The generic pipeline is instanced each time in a different way by a wide range of approaches for each of the fields of application. [29] Have made an attempt to categorize various object tracking approaches. Therein, they followed a basic scheme that has been introduced by [43] before. They have split the problem of object tracking in a first stage by the form of object representation and they distinguish be- tween shape based and appearance based representation. According to [32] and [26], shape based representation includes models such as object contours, skele- tal models, geometric shapes and points. Object contours thereby refer to the boundary of an object with the object silhouette inside that region. Together with skeletal representations which are area reduced versions of contours, they are suitable for tracking complex non rigid objects. Point based representations mostly rely on the centroids of objects and therefore can be used only to track relatively small objects. Related tracking methods therefore could be based on techniques such as Kalman filtering [34], particle filtering or active contour methods [14, 30]. Furthermore, [29] and [26] mention probability densities, templates, active appearance models and Multi view appearance models among the appearance-based object representation. Whereas probability densities comprise only parametric densities, nonparametric densities or histogram information, templates add geometric shape or contour information of a single view. Therefore these models are suit- able only for non-deformable objects whose appearance does not change considerably during tracking. Active appearance models and multi view active appearance models need a training phase for simultaneous modeling of the object shape and appearance in a single view or

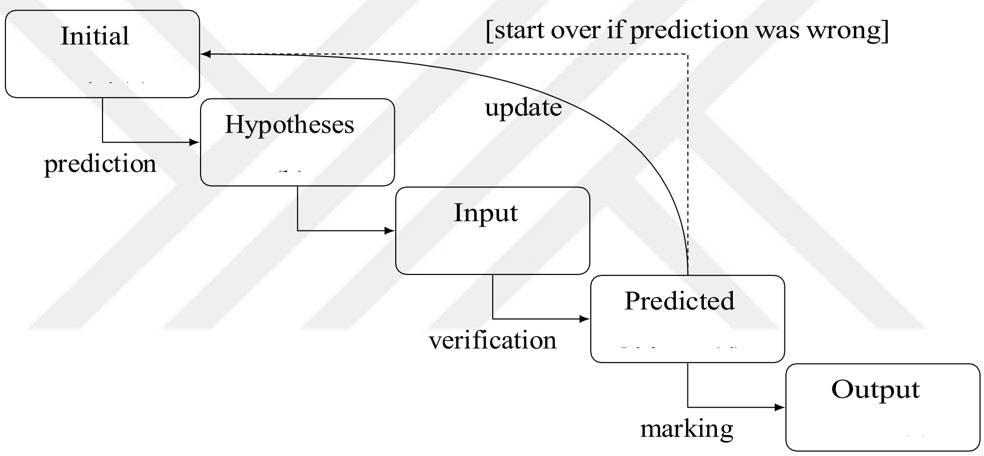


Figure 2.3: Generic top-down object tracking pipeline, showing the processing steps between the initial model (a) and the output frame (e). Therein, hypotheses (b) are being generated and applied to the input frames (c) in order to predict objects and to verify them (d). Finally, the objects are being marked in output frames (e).

Multiple views. Associated tracking methods comprise sparse variants such as based on point descriptors (e.g. SIFT, SURF, KLT) [6,] or dense manifestations such as histogram-based trackers (e.g. mean shift, CAMSHIFT) [17, 10, 25]. In Figure 2.3 the basic attempt for a categorization is visualized in tree form [34,]. Nevertheless, this attempt is not the only way of a categorization of object tracking approaches. [15] for example distinguish between deterministic and stochastic methods. The different approaches to object tracking comprise pipelines that include various filters which are being applied to one or more of the frames fi ∈ F-input or to outcome of intermediate steps of the pipeline such as oi ∈ Oseg. Those filters can be of various type such as image processing filters or machine learning algorithms.

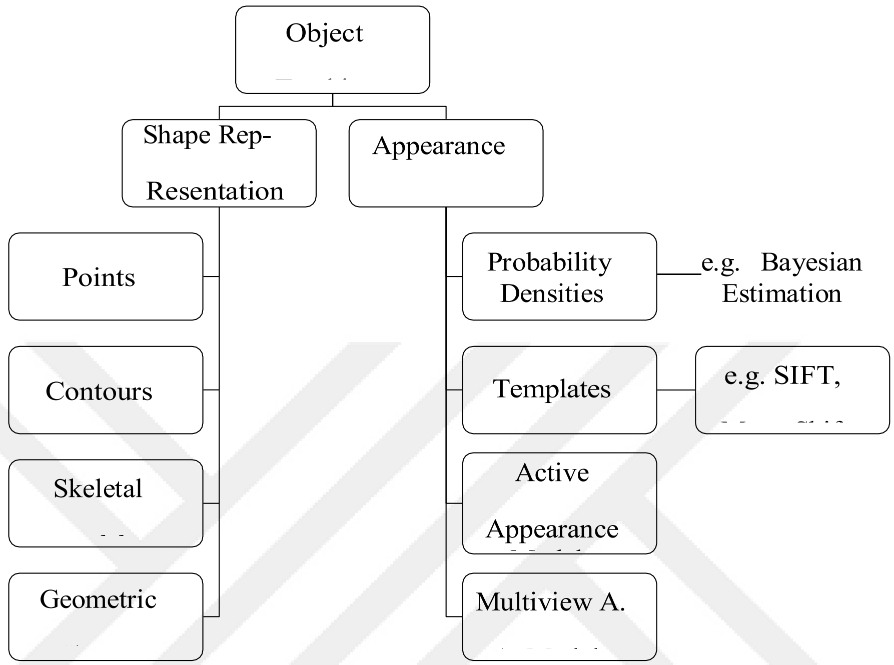


Figure 2.4: Categorization of object tracking techniques

Especially if pre-processing is required, image processing algorithms are involved. Authors such as [35] categorize image processing algorithms into global and local applicable filters. In global image processing thereby every output value is dependent on all values of the image, thus comprising algorithms such as histogram generation or mean value computation. Thus, the algorithm generates one output value that depends on all pixels of the image. Considering local image processing methods, every output value gi0 depends only on a single pixel gi or on a neighborhood, e.g. N8 (gi). Local image processing comprises algorithms such as an intensity changes or mean filtering. A change of the image intensity by a value of Δ can be defined as

$$\forall gi \in V:\ \ \phi f\,(gi0) = \phi f\,(gi) + \Delta$$

The three different top-down approaches to object tracking that are being used throughout this thesis as examples of object tracking pipelines in contrast to the bottom-up approach. Since there is no general way of a categorization and since there is a vast amount of different approaches to the subject of object tracking, this part of the thesis only considers three approaches that have been popular in terms of publications over the last years. The chosen approaches have also been used in the field of ultrasound video streams already and may therefore be suitable for the given problem of nephrolith tracking. The following subsections introduce the approaches of mean-shift tracking, particle based tracking and active contour tracking, focusing especially on the pipeline models behind.

2.1.2 Region Growing

Object tracking in US video streams has been continuously researched, as has been shown by various authors [11] in the past few years. In [33] presented a model based robust approach for shape tracking in US images. Carneiro et al. [13] proposed a method for automatic detection and measurement of fetal anatomies in the same year. However, object tracking is not only an issue in medical image processing but instead comprises a wide range of applications. According to [12], some of the most popular object tracking applications include air space monitoring, video surveillance, weather monitoring and object tracking. Air space monitoring is used to keep track of flying objects in controlled air spaces for security reasons or national defense. Weather monitoring may be used to generate forecasts. Video surveillance and object tracking are related in the security sector but may also be used for e.g. manufacturing processes. Additionally, the military industry provides purpose for object tracking, e.g. for targeting systems. In [17] have presented a survey about object tracking, comprising a list of applications. According to them, object tracking is used for human identification, automated surveillance, video indexing, human-computer interaction, traffic monitoring and vehicle navigation. Systems for human identification may identify persons by reference to gait or facial features and could be used for enterprise security or airport surveillance. Video indexing could be utilized for automated annotation and search of videos in large database systems. Gesture recognition and eye gaze tracking are both subcategories of human-computer interaction with state-of-the-art applications in the sector of video games. Traffic monitoring systems may be utilized for congestion prevention or police actions. Vehicle navigation keeps track of traffic signs and the course of the road and may be used to warn drivers about dangerous conditions. Nevertheless, Challa and Yilmaz et al. do not claim to provide complete lists but try to give an insight into the large variety of applications. Beside popular applications, there is also literature on specialized applications such as presented by [23] involving object tracking in geo-coordinates [45]. All the mentioned applications provide for different scenarios for the object tracking systems. Thereby, particles which represent random values of a state-space variable are being generated, used as a description in a model of the tracker and finally weighted to determine their likelihood of being a representative particle. Therefore, a set of particles approximately describes a PDF of the target. Since that PDF approximation can be used in other tracking approaches such as in the mean-shift method as an alternative to histogram-based target models, particle filter based tracking is not only a standalone tracking approach [31].

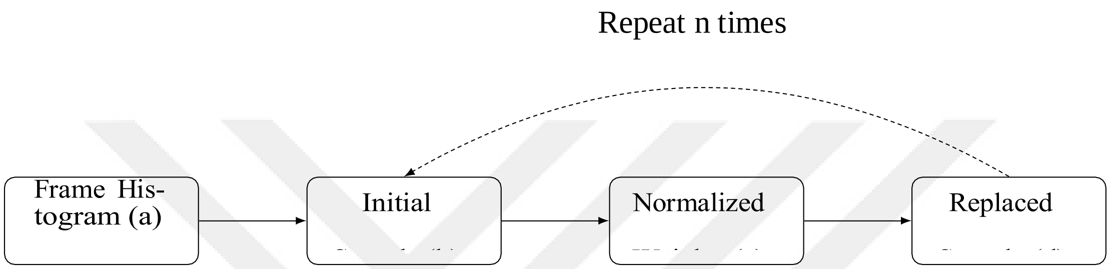


Figure 2.5: Sequential Monte Carlo pipeline according to the algorithm of [41] as an initialization for the tracker. An initial distribution (a), e.g. frame histogram, as well as an importance distribution (b), e.g. feature map of a template model, are being sampled. Subsequently, weights (c) are being calculated and normalized before the initial distribution is being resampled using the weights (d).

Kroese [31] has defined a sequential Monte Carlo algorithm in a generic form that samples approximately from a given density function. It consists of four major steps: initialization, importance sampling, selection and a stopping condition. There are also more advanced versions of such a particle filter but this simple form is sufficient for showing the pipeline behind this approach. The initialization step comprises a, possibly random, sampling of a measurement vector from the initial distribution such as the frame histogram. In the next step, a second vector is sampled from an importance distribution and weights are being calculated as well as normalized. The importance sampling is a critical step in the pipeline, leading to more particles being sampled in important regions. However, biasedness of the estimator is being preserved by including the importance distribution not directly but in the form of normalized weights. Importance sampling theory has been discussed extensively in the literature, e.g. [41]. A practical implementation from M. Couprie and G. Bertrand [19] for object tracking utilizes a feature map of a 3D human object model as an importance distribution. Using this weighted sample, replacement particles are then being sampled from the initial distribution. Next, the whole process is being repeated beginning from the second step, until a specific number of iterations has been made. A simple pipeline that represents the initialization part of the tracker visualizes this algorithm [29] have presented a combination of a particle filter and an active appearance based model for a real-time multiple object tracking system.

2.1.3 Statically Region Merging

In statistical region merging, they have utilized the massively parallel computation power of GPU in order to realize the particle filter. Thus, they have shown that a particle filter can be designed efficiently multithreaded. They also state that particle filters require high arithmetic throughput and have low communication and storage costs, which makes them suitable for parallel com- potation. Their tracker comprises three steps: an initialization phase, the tracking itself and a subsequent output generation. The initialization stage is performed continuously in a separate thread on the CPU. It is responsible for scanning the incoming frames for new objects, utilizing a Viola and Jones boosting algorithm [30] for detection. It also extracts a sparse description template that is used as initial model in the tracking stage. The tracking stage itself comprises the particle filter and runs partially on the CPU and on the GPU. Lozano and Otsuka [49] claim that only the weighting part of the algorithm has been moved to the GPU because it is the computationally most expensive part. The output generation phase eventually comprises an averaging of the best results of the tracker before marking the output. A visualization of this implementation is given in Figure 2.6

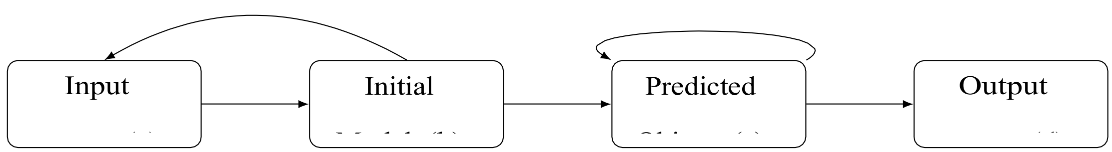


Figure 2.6: Particle filter-based tracking pipeline. Input frames (a) are being continuously searched for new objects and models (b) are being extracted to initialize continuous tracking of those objects (c) before averaging the results and marking them in output frames (d).

[18] Have presented another application of particle filters for object tracking. Instead of an active appearance representation such as has been used by Y. Lin, Q. Yu, and G. Medioni [39], they proposed a combination of a particle filter and the Bhat- tacharyya coefficient [16] which is commonly being used in mean-shift approaches. They claimed to have brought together efficient tracking components. Following that, [20] proposed a combination of a particle filter and the whole mean-shift algorithm. Instead of an improved performance however, they have searched for a more

robust tracking solution. This approach includes tracking with the mean-shift method and subsequent avoidance of local maxima by using a small number of samples of a particle filter.

2.1.4 Heuristic Methods

To obtain a good perspective information of the scenario, a vanishing point-based perspective estimation is used. In road scenarios, lane markings and guard rails are parallel to each other, and they will pass the vanishing point on the image plane. Also, a roughly plain surobject of the road is assumed. Single image vanishing point estimation is implemented based on [28] and updating vanishing point position is done periodically instead of every frame to save computational load. Using of vanishing point in the image, the plain surobject could be located on the image plane. And the perspective information of the objects on this surobject would be inferred. Based on this perspective geometrical information, the scaling is performed as examples shown in the right column in Figure 2.7, that the small objects will not be cropped as illustrated in the middle column. The number of scales and the size of each scale can be determined through statistical analysis of the performance of each network structures on a given dataset and choosing best of them. However, this trivial process involves a lot of retraining and testing that requires plenty of computational resources and time, to simplify the process and demonstrate the effectiveness of the proposed perspective aware strategy, heuristic numbers are used as detailed in the following experiment section Recent advancement of parallel computation hardware and explosively generated visual data on the Internet boost the neural networks into deeper structures. For the ANNs commonly used for vision tasks, with deeper structures or layers, the networks are capable of learning better representation of various object and utilizing large amount of data, usually causing over fitting problem in shallowing learning algorithms. Like the traditional detection algorithms such as Haar/AdaBoost, HOG/SVM or DPM, the CNN based detectors perform the same tasks of detecting interested objects in the images and output the location. A Regional Proposal Networks (RPN) [18] based algorithms is employed in this paper. The RPN is the proposal network that generates anchors at different scale with different width/height ratio. The base network before the RPN can be different types of networks, such as VGG16/VGG19and Residual Networks (ResNet) [37]. Comparing to the traditional ANN detectors that use ANN as a classifier to classify object image patches proposed by an external proposal algorithm. This method moves the proposal module into the neural network and trains an end to end solution for object detection with real-time or near real-time (depends on different base networks) performance on a GPU.

Figure 2.7: Common scaling strategies in traditional detection algorithm (left column) and the proposed perspective aware scaling strategy (right column)

Object detection algorithm introduced in aforementioned section uses RPN to generate proposals of potential object location. This type of detection algorithm performs poorly on small targets in road scenarios. One potential improvement is to adopt the multi-scale strategy (left column in Figure 2.7) as used in transitional object detection algorithms, by building a pyramid of images resized from original image sizes to detect the objects and then convert back into the original image scale. However, scaling may not help detect small objects far away since most of the scaling strategy is based on center point of the image plane, while small object may not always locate in the center as illustrated in Figure 2.7. The middle column shows scaling using center point may cut out the interested small targets that are supposed to be detected. To make better use of this strategy, a perspective aware scaling is proposed.

2.1.5 Artificial Neural Network

Deep learning is a branch of machine learning based on neural networks. It could be supervised, semi-supervised or unsupervised. Supervised learning is the one that learns labelled training data as an example and maps an output based on the level of learning. The training data is analyzed, and an inferred function is produced that could be useful for generating newer examples. In ideal terms, a supervised algorithm can accurately determine class labels for unobserved instances. Similarly, semi-supervised learning is a class of machine learning that uses both labelled and unlabeled datasets for training purposes. However, the amounts of unlabeled training datasets are far more compared to the labelled set of training data. It has been observed that the learning accuracy is considerably improved when larger unlabeled set of data is used with smaller set of labelled data. Likewise, unsupervised learning is a machine learning technique that figures out unidentified outlines of dataset without any pre-existing labels. This type of learning mainly uses the principal component and cluster analysis. For Principal Component Analysis, an orthogonal transformation is used in order to alter a group of observations of correlated variables to linearly uncorrelated variables, also referred to as principal components. Similarly, for cluster analysis, the data that has not been labelled is grouped and is further clustered or categorized to evaluate common ground in datasets. This in fact, helps detect anomalous objects that do not appear familiar or do not fit into the group. Basically, deep neural networks are in fact artificial neural networks with multiple hidden layers in between the input and output layers. Due to the presence of multiple layers, deep neural networks are capable of modelling complex non-linear relationships with complicated datasets which most similarly, performing single-layered shallow networks fail to do. The main advantage of working with deep learning algorithms is that they do not need structured or labelled data to perform prediction [18]. Deep neural networks rely on the hidden layers of artificial neural networks and learns through each layer hierarchically. The neurons of deep learning are classified into 3 different categories:

1. Input Layer
2. Hidden Layer
3. Output Layer

Figure 2.8 shows the graphical representation of deep neural networks. The input data is received through the input layer where it transfers the data to the first hidden layer. In Figure, we have altogether 4 different hidden layers. All the mathematical computation occurs inside the hidden layers. The data is computed in each first hidden layer and its output serves as the input of the second hidden layer and so on. It is therefore, important to understand what number of hidden layers and the number of neurons in each of these hidden layers would help get the best overall prediction. Similarly, after successive computations in each of the 4 hidden layers as shown in figure 2.8, the output layer returns the final output data, which we call the predicted value.

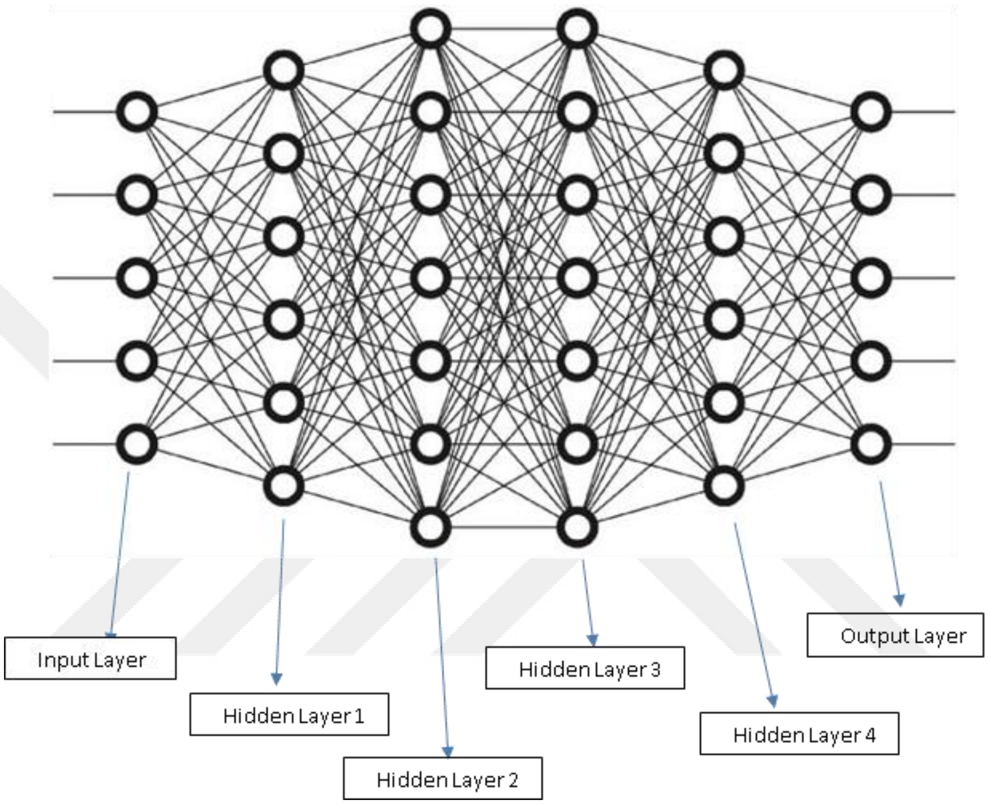


Figure 2.8: Artificial neural network

An image is passed through an artificial neural network and assigned importance based on learnable weights and slight biases to certain features or items in the image that can differentiate one part of the object from another in same or dissimilar images. The preprocessing time required in deep convolutional neural networks is comparatively lower in comparison to other classification algorithms. Convolutional Neural Networks can purposefully seize both spatial and temporal dependencies throughout the image by applying numerous relevant filters. Especially, with the advent of GPU and improved algorithms, deep neural networks usage has surged several folds for analyzing visual imagery. Some of the algorithms used in the current study are described as follows.

2.2 OBJECT TRACKING

Object tracking has a long history in both the radar and computer vision research communities. The fundamental task of object tracking is sequentially estimating the states of interested objects given observations or measurements from sensors in a scene or multiple scenarios. The states of targets depend on the applications, and typical states include position, velocity, and acceleration. A sequence of states formats the spatial and temporal description of the interested targets. Based on different types of sensors, various algorithms have been proposed to track multiple types of objects Object tracking has a long history in both the radar and computer vision research communities. The fundamental task of object tracking is sequentially estimating the states of interested objects given observations or measurements from sensors in a scene or multiple scenarios. The states of targets depend on the applications, and typical states include position, velocity, and acceleration. A sequence of states formats the spatial and temporal description of the interested targets. Based on different types of sensors, various algorithms have been proposed to track multiple types of objects

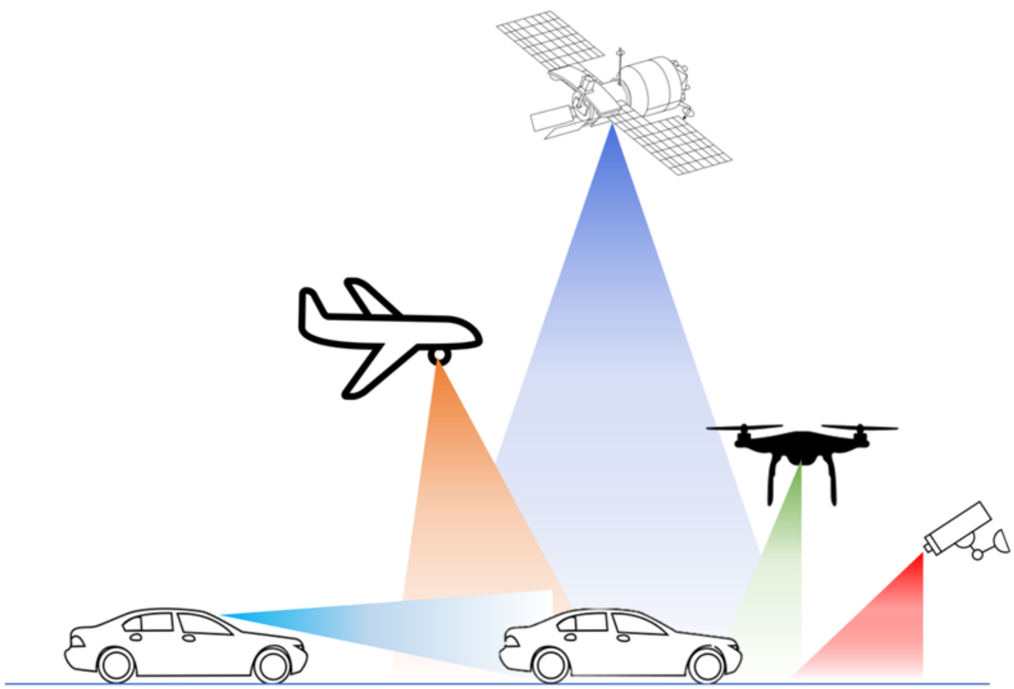


Figure 2.9: Different types of ground target tracking platforms

Multiple types of platforms could be employed to track ground targets. For example, as shown in Figure 2.9, common platforms include satellites, aircrafts, traffic surveillance systems, and automotive safety sensory platforms. Depending on the distance from sensors to the objects of interest to be tracked, the platforms could be divided into two categories

Long-distance sensory platforms Long distance surveillance platforms are usually mounted with sensors like long range Radio Detection and Ranging or Radio Direction and Ranging (radar), such as Ground Moving Target Indicator (GMTI) radar, or Electro-optical/Infrared (EO/IR) systems. These platforms cover relatively large surveillance area and handle tracking problem with large number of targets. Due to the long-distance from sensors to targets, objects to be tracked are often modeled as point targets in radars, or small image patches with vague details in images, and the maneuverable movement range of the target is relatively small within the region of sensor's field of view (FOV) compared to the size of the FOV.

Short-distance sensory platforms Short distance platforms are often referred as cameras, mid-range/short-range radars, or Light Detection and Ranging (LiDAR). For example, one of the common short-distance platforms are traffic cameras. Com- pared to long-distance sensors, these platforms have smaller range and area of FOV, and less number of targets to be tracked, however, more details of the object can be observed. For example, in radar/LiDAR measurements, multiple reflections that originated from the same target would be received. In image platform, the objects would occupy more pixels, meaning more details are captured for the object. Rich features observed from these sensors would increase the accuracy of tracking and reduce errors. Before elaborating the proposed algorithms for ground target tracking, a brief review on general object tracking theory is presented in the following section as the fundamental basis for the derived algorithms.

2.2.1 Video Based Objects

Video based objects can be evaluated in the context of autonomous vehicles in order to reduce the risk of collision. To pose timing requirements of the object detection pipeline, simulations of stopping distances for a vehicle were done. By calculating a maximum allowed latency, it could be used as a reference when reasoning about the speed of the algorithms. Hardware requirements was defined based on the timing requirements posed from the stopping distance simulations.

YOLOv3 was initially implemented in order to evaluate its baseline performance. Following the implementation of YOLOv3, a first implementation of the object detection methods was planned. After this, videos could be rendered out to run through the YOLOv3 algorithm, and these gradient based learning allowing testing and evaluation. Activation functions sit between layers in a network to introduce a non- linearity. Otherwise, the operations could be regenerated to a simple linear transformation and remove the benefits of building a deep model. The rectified linear unit or ReLU which is the recommendation in the activation function for neural network. It's defined as the maximum of 0 and the input value and can be described as a nonlinear function made up of two linear pieces. Because of this, it preserves many of the properties that make models easy to optimize and generalize well on new data. Since the introduction of ReLU other variants have built on its success like PReLU , LReLU and ELU . A lot of research does not only concentrate on one single characteristic, it heads for combinations of biometric tokens. In this work we are concerned with the object detection and object analysis token and we focus on eyeglasses detection and segmentation as a pre-processing step for object detection and analysis applications. For robust object detection it is essential to have normalized pictures, which fulfill the ICAO requirements for Normalized object images should be interference-free and should contain a frontal pose, neutral expression and 'normal' eyeglasses with a typical frame, clear glasses and no reflections. Therefore it is necessary to detect eyeglasses and - if present - to delineate and evaluate the eyeglasses. This leads to solving problems in the areas of classification and segmentation. Normalized images which come up with all the necessary qualifications increase the reliability of object detection and detection software. The task in this work is to automatically detect the presence of eyeglasses and to locate the eyeglasses in the object portrait image [15]. With the located eyeglasses several different parameters of the eyeglasses appearance can be calculated. All these parameters lead to decisions, whether an object image containing eyeglasses is suitable for a 'Machine Readable Travel Document' or not. Magnetic Resonance Spectroscopy (MRS) measures the spectral peaks of different compounds containing protons in the tissue of interest. Instead of producing an image, MRS data are usually reported as spectra of different metabolites. Nonetheless, it is possible to integrate under the metabolite peak of interest and generate a low resolution metabolite image as well. Studies have shown that an increased choline peak is indicative of cancerous lesions due to increased cellularity, and MRS can be useful for diagnosing lesions greater than 1 cm in diameter. High field magnets offer two advantages for mrs: the frequency separation between peaks is greater at higher fields making it easier to differentiate between compounds and the increased snr at high fields enables greater signal for metabolite peaks which exist at much lower concentrations than water in the body.

2.2.2 Image Based Objects

The Placement of camera sensors is as important as having good image quality in image-based objects. Poor placement could affect the field of view and leave unsupervised areas. Cameras could be placed in many different ways and research suggesting a specific positioning was not found. Different car manufacturers place their cameras in different places and use different kinds of focal lengths with their lenses. Cameras could be placed in front of vehicles, at the sides, behind the wind shield, in the rear or on the roof of the vehicle. Wide angle lenses would solve problems with poor field of view but also compromise detail because of lens distortion [10]. Using a large amount of cameras could also be a solution, however would that be more expensive and more complex in computation. With the increase of computational power from CPU's and GPU's it has now been possible to deploy machine learning and deep learning algorithms to vehicles in order to increase their levels of autonomy. Algorithms within regression, pattern recognition, clustering and decision matrices are sections of machine learning that could be used within autonomous driving. Deep learning, a form of artificial neural networks but with multiple hidden layers, is the most commonly used in self driving cars since the levels of complexity is able to increase tremendously. Neural networks have been endorsed by Tesla and Waymo and is able to detect objects, make decisions and much more. These questions will be explored for detection of glasses in facial images by implementing an example IoT use case up to the level of a working prototype, in addition to literature research. These questions could also be explored exclusively through literature re-search, so why go through the effort of creating this prototype? We feel that many issues which might be otherwise missed or considered trivial will be re- veiled this way. By going through the entire process from the ground up we will by necessity experience the design decisions that have to be taken at each step of the way, and think about their security impact. Furthermore a running example serves to clearly illustrate many of the security issues to tackle. The use case that we will be exploring is the tracking of a passenger's bag- gage after checking it in at an airport. In the system a passenger can use his smartphone to track when his bag has passed certain scanners in the system, as a toy implementation. A cheap passive RFID-tag will be attached to every bag entering the system, which can then be picked up by scanners set further along the track for image-based objects. Perfusion measures the ability of blood flow into a specific tissue, and predominantly, most studies have demonstrated contrast-agent based perfusion [14]. There are several MR techniques for measuring perfusion in the breast with or without a contrast agent.

2.2.3 Object Tracking by Energy

The object tracking by energy and all these parameters lead to decisions, whether an object image segmentation and tracking is suitable for a ’Machine Readable Travel Document’ or not. Magnetic Resonance Spectroscopy (MRS) measures the spectral peaks of different compounds containing protons in the tissue of interest. Instead of producing an image, MRS data are usually reported as spectra of different metabolites. Nonetheless, it is possible to integrate under the metabolite peak of interest and generate a low-resolution metabolite image as well. Studies have shown that an increased choline peak is indicative of cancerous lesions due to increased cellularity, and MRS can be useful for diagnosing lesions greater than 1 cm in diameter. High field magnets offer two advantages for mrs: the frequency separation between peaks is greater at higher fields making it easier to differentiate between compounds and the increased snr at high fields enables greater signal for metabolite peaks which exist at much lower concentrations than water in the body. However, mrs is difficult to use for diagnosis because it is not always volumetric, and correcting for the b0 homogeneities through techniques such as shimming can be technically difficult. Magnetic resonance elastography (mre) is the closest image processing technique that resembles the physical palpation exam, and it measures the viscoelastic properties of the tissue. It is a phase-contrast image processing imaging technique that visualizes and measures acoustic strain waves in tissues after harmonic mechanical excitation. The shear modulus of the tissues can be calculated from the collected data, and various studies from breast tissue specimens have shown that certain carcinomas have higher shear modulus than normal fat and glandular tissue. [27] Developed an eyeglasses detection algorithm, where they analyze the nosepiece in the object image. They did several experiments and demonstrated that the area between the two eyes is the most important criterion to determine if glasses exist or not. After calculating several features of this area, the presence of glasses is determined. The evaluation database contains 419 object images, where 151 people wear eyeglasses. All the object images are frontal view and the eyeglasses have various shapes and colors. The correct detection rate on their database of their classifier is 99.52 %. With all these advances, image processing tools are useful for parsing the abundance of information from a breast image processing examination. Perhaps, the most widespread of these tools is computer aided detection (cad) software for analyzing and displaying kinetic data. the cad software produces a 3d color map of the breasts, where the color displays lesions that may be suspicious based on kinetic data such as upslope, percent enhancement, and delayed enhancement . These systems have the potential to improve breast image processing efficiency and reduce the number of false-positives, but they are highly susceptible to motion artifacts and noise. For severe motion artifacts, it would be beneficial to perform image registration before passing the image data into a cad program. Classification of lesion morphology through computer tools has promise, and these novel techniques would require extensive validation before implementation into a clinical protocol. Improvements in dedicated breast image processing coils have been the single most important advance in achieving high spatial and temporal resolution of the breast through parallel imaging. Due to the fact that readout is usually performed in the anterior/posterior direction to limit impact of cardiac motion artifacts, this leaves the superior/inferior and left/right directions available for applying parallel imaging. the high density breast coil (ge healthcare, waukesha, wi) and similar volumetric commercially-available coils have made a substantial difference to breast imaging, allowing at least 2× parallel imaging acceleration and marked snr benefits compared to the body coil or a single large rf coil. However, there is still room for improvement in terms of snr and parallel imaging with breast coil arrays. Sentinel (now a subsidiary of hologic) was one of the first commercial companies to address this issue in their design of biopsy-compatible, 8-channel and 16-channel breast coils with adjustable side plates for bidirectional parallel imaging. In this dissertation, we will detail how we designed, developed and demonstrated an 18-channel closely fitted coil for clinical bidirectional parallel imaging of the 3d image processing.

2.2.4 Object Tracking by Contour Model

Active contour-based tracking falls into the category of contour based object representation regarding the categorization. Active contour models or snakes haven been mentioned first by E. Le Pennec and S. Mallat [38]. They have used energy minimization as a framework for designing low level energy functions which provide in their local minima a set of alternative solutions available to high level image processing. Snakes are commonly being used for edge detection, line detection and contour detection as well as for motion tracking of those objects. Gunn and

Nixon [30] claim that snakes incorporate a global continuity and curvature based edge detection combined with a local edge strength feature. Leymarie [44] describes snakes as deformable curves or contours that are composed of elastic materials making them resistant to stretching and bending. According to [36], snakes represent a general technique to match a deformable model to a given frame by minimizing its energy function. Book applications for sparse image processing comprising for example image denoising, image compression, deblurring or restoration tasks. These procedures may be interesting for object tracking in terms of preprocessing of the incoming frames. Especially when tracking objects in US images, denoising may be a critical process in order to increase the signal to noise ratio (SNR). However, some of the implementations of sparse image denoising are only capable of removing additive noise. Therefore, the usefulness of sparse processing for tracking has to be decided for each particular case. [37] Have used sparse processing for image segmentation. They have presented a multi-object level set method that is computationally efficient and is based on a geodesic active contour representation. According to them, the time complexity of the algorithm is linearly growing with the number of pixels in the images but independent of the number of segmented objects. D. P. Kroese [34] has presented a level set approach for reconstructing shapes of 3D objects. The proposed level set approach is based on a sparse-field algorithm which tracks active pixels in a set of linked lists. According to [24] the time complexity of this algorithm is less than the complexity of the original level set approach. The presented sparse processing approach is based on an information preserving internal representation of the given images. Particle injection based approaches, on the other hand use an extremely sparse representation of only the most important parts of the frame. These representations reject a major part of the information comprised in a frame. Nevertheless, this often allows for fast and robust tracking and avoids occlusion effects when tracking multiple objects as has been shown in the past in Particle injection is frequently used for top-down object tracking approaches but can also be utilized for bottom-up trackers. Chowdhury et al. [16] have proposed a cell tracker based on bipartite graph matching techniques. They assume that the cell objects in each frame have been segmented already and utilize cell centers as vertices of a bipartite graph. The graph is is being utilized for matching cells in subsequent frames which runs in polynomial time. The preceding segmentation could possibly be replaced by Monte Carlo particle injection, resulting in a tracker that is completely based on sparse processing. However, only a small amount of literature seems to exist about a combination of these techniques for object tracking.

3. METHODOLOGY

The purpose of this study was to develop a quantitative experimental prototype for object segmentation and tracking in digital image using Active Contour Model and Heuristic Method. The computer ran Ubuntu 16.04 with the NVIDIA CUDA and cuDNN libraries installed to take advantage of the GPU. As the primary programming language Python was chosen due to its simple syntax and popularity in the deep learning field. The code was briefly tested with Python as well and seemed to work. Keras was used as the framework for training models because its high-level syntax allows fast prototyping and testing. The development environment was a Jupyter Notebook, which allowed a flexible execution of code snippets instead of running the entire program for every single change. The methodology adopted for implementing an object segmentation and tracking in digital image. The main components consist of first, a GPU-enabled backend (on premise) which is designed to ensure that very deep models can be trained quickly and implemented on a wide array of cameras in near real time. At the heart of the proposed AI-enabled system is the development and training of several deep convolutional neural network models that are capable of detecting and classifying different object segmentation and tracking in digital image into its constituent objects. . Manually annotated objects served as the primary source of dataset used for training these models. To enable the system to be situational aware, different object tracking algorithms are implemented to generate trajectories for each detected object on the tracking at all times. The preceding steps are then combined to extract different object flow variables and monitor different object conditions such as queuing, crashes and other object scene anomalies. The AI-enabled object monitoring system is capable of tracking different classes of vehicles using YOLOv3, spotting and detecting classification and tracking object segmentation in real-time. The following sections will describe the details of each of the systems components including the algorithms deployed in this study.

Figure 3.1: Common scaling strategies in traditional detection algorithm and the proposed perspective aware scaling strategy.

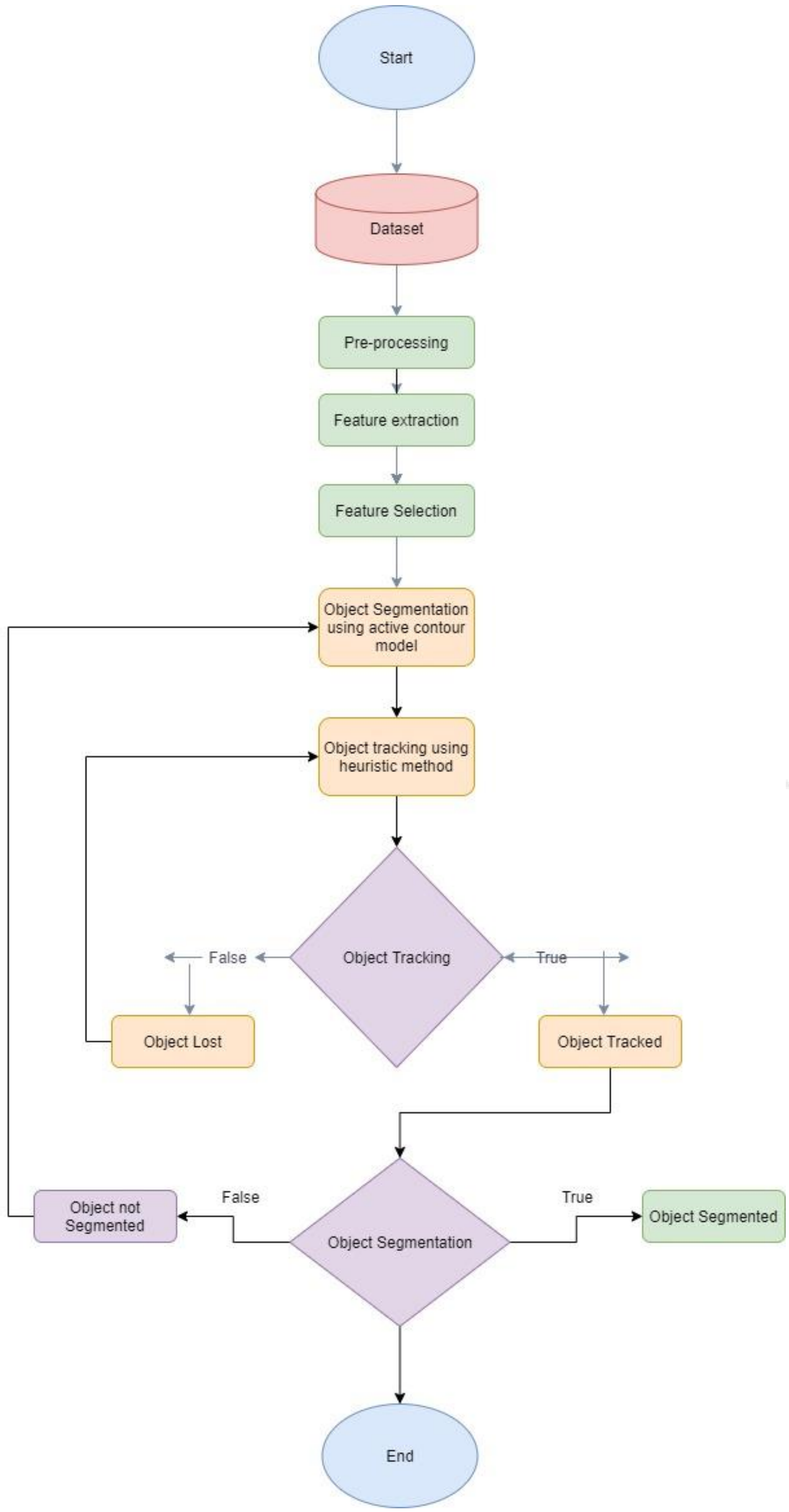


Figure 3.2: Process Flowchart

3.1 MULTI-CLASS DATA ASSOCIATION

In most of the traditional object detection algorithms as mentioned above, only one type of objects could be detected. In order to detect multi-class or multiple types of objects, the models for each different type of objects need to be evaluated one at a time, which will significantly increase the detection processing time. Thus, the detection part will limit the capability of tracking multiple different types of targets in videos. However, one of the benefits from adopting the CNN based detectors is that different type of targets could be detected with only one pass evaluation of the model. Therefore, multiple different types of object could be detected without increasing any computational load by using CNN based detectors. To track multiple targets with different class information, one could either use an object class agnostic way or class specific way to handle feed the data association. The former method treats all the detections as type agnostic targets and tracks them indistinctively. The second method treats targets separately and conducts data association for each separate class of objects by solving multiple 2D assignment optimizations with subsets of association pairs. Although the second method may increase the difficulties in designing the system for multiple classes, there are also benefits of associating detections and tracks for each class separately. The first benefit is the reduction in association computational load. Building an association pairs of detections and tracks of all the types of targets would result in a huge association matrix therefore increasing unnecessary computations. With the availability of class information, the association matrix could be divided into blocks and the association could be handled individually. Secondly, with the class information, inter-class association error could be avoided. However, this does introduce an extra step for handling occlusion, since only considering single class of object would blind that class's tracking algorithm about occlusions between different types of objects. Thirdly, based on class specific association algorithms, separate tracking algorithm including different track management methods could be utilized to maximizing the performance for tracking multiple types of targets.

3.2 HIERARCHICAL OBJECT CLASS TREE FOR MULTI-CLASS DATA ASSOCIATION

One drawback of the aforementioned multi-class data association algorithm is that the object class is not always classified correctly by the detector, especially when the object is being partial occluded. Also, the same object may have different class labels that cause missing detections for some targets and broken tracks. A hierarchical object class tree is proposed to handle this problem.

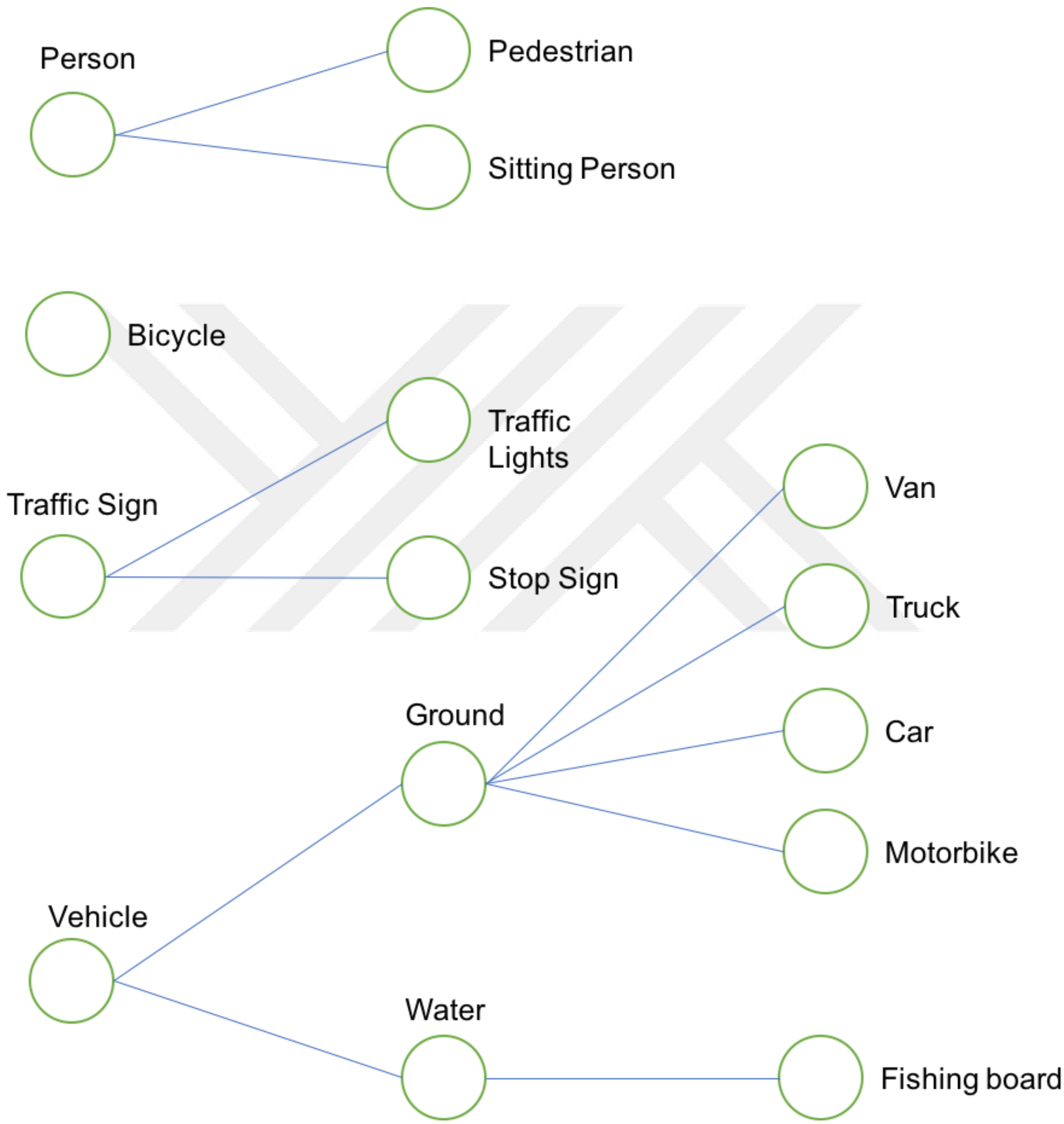


Figure 3.3: An example of hierarchical object class tree representation based on their common attributes.

An example of hierarchical object class tree is given in Figure 3.3. In the detection framework, only the end leaf nodes are used in the last softmax loss function to predict class information, their relationship is not addressed. Handling misclassification problems in the detections, based on the hierarchical representation, depends on the classes that are similar to each other and prone to have misclassification error, the division of subsets based on classes is modified to include some of their parent nodes. For example, trucks, vans and cars could share a same subset, same for people and bicycle. How to build the hierarchical tree of different classes depends on available classes in different training datasets, and the principal is to include visual similar classes and classes that are potentially incorrectly classified.

3.3 DATASET

VTB-100 s a public dataset collection that contains 100 sequences of images captured in different lighting and camera conditions. The datasets consist of different types of objects to be tracked, such as pedestrian, birds, face, toy, vehicle, and etc. Variety of the objects, rich conditions of the environment, and different types of camera performance and camera motion create different problems for the trackers to handle, and therefore VTB-100 is a reasonable testbed for benchmarking the proposed tracking algorithm. To characterize and summarize the challenges in tracking objects in the datasets, the authors proposed several different attributes as listed in Table 3.1.

Table 3.1: Attributes of VTB-100 datasets representing challenges for tracking.

Attributes	Description
Illumination Variation (IV)	Significant change in illumination condition on the target.
Scale Variation (SV)	Target size varies in different frames.
Occlusion(OCC)	Full or partial occlusion on the target.
Deformation(DEF)	Target deformation.
Motion Blur (MB)	Blurred target appearance due to motion.
Fast Motion (FM)	The movement of the target is larger than tm pixels (tm=20).
In-Plane Rotation (IPR)	The target rotates within the image plane.
Out-of-Plane Rotation (OPR)	The target rotates outside the image plane.
Out-of-View (OV)	The target is partially out of field of view.
Low Resolution (LR)	The area of the target's bounding box is less than tr pixels (tr =400).

3.3.1 Training, Validation, and Testing Sets

All of the data is divided for the different purposes in this segment in which the portion of the data for the training will draw direct conclusions from the validation data. But if the accuracy of the validation set drops below the results of the training data of object tracking, the network is starting to memorize the data it knows rather than generalizing on the concept. This subset that will go through the process of testing is called testing data. In contrast to the validation set, it's only used once in the very end, to give a final score. The idea is that by building a model based on the validation results a certain amount of information bleed occurs, where the network will implicitly learn from the validation data. In order to avoid bad results, testing data is normally used as a reference for the performance.

- Training Set: 74% of the data
- Validation Set: 13% of the data
- Test Set: 13% of the data

3.4 TRACK MANAGEMENT

Track management is an important component in a multi-target tracking algorithm. It aims to manage states of the tracks, such as the creation of new tracks and the deletion of dead tracks. In radar-like sensor based tracking, adding a confirmation step instead of creating tracks immediately is for reducing number of false alarms, caused by random noise, being treated as targets. However, this may not be effective in other sensor applications in which false alarms are not caused by random noise, like in image tracking, a false alarm caused by miss-classification may not always randomly show up in detections, in other words, they can be detected like a target continuously in several consecutive frames. This difference requires new approaches to handle the track management in visual tracking.

On a moving platform, due to the relative motion between the camera base plat- form (the ego-vehicle) and other objects, object showing and disappearing happens almost all the time. Therefore, the quality of the track management will considerably impact the entire tracking performance, like false alarms rate, missing tracks, track identity switch, and track fragments. The track states are often defined as,

1. Tentative tracks. A tentative track is the initial state when a track is created, and it remains tentative until it is being confirmed.

2. Confirmed tracks. A confirmed track means the track has enough detections and high detection confidence to be confirmed as a true object track, otherwise it will be tentative. A confirmed track will be outputted as tracking result while tentative tracks are only maintained internally.

3. Dead tracks. A track will be marked as dead track if it is considered as out of the scope of tracking scenario.

Track management algorithm manages the above states by creating new tracks, up- dating tentative states, and eliminate dead tracks. The track creation and deletion determines the number of missing tracks and false tracks; a good management algorithm should minimize these numbers.

3.5 TRACK CREATION

Track creation initializes new targets and adds them into the existing tracks for maintenance. After the data association, the measurements left unassociated will be considered for creating a new track. Two criteria are examined to determine if the unassociated measurement is a new target, one is the detection confidence. Since the detector is trained based on manual labeled data, higher confidence means more likely the object is the target. A target is created immediately when the unassociated measurement's detection confidence exceeds certain threshold. For the rest lower threshold measurements, a tentative track is created and normal tracking is performed there- after except tentative track is not outputted. The tentative tracks will be upgraded to confirmed tracks only if the first criteria are satisfied or consecutively associated with lower threshold detections for a certain number of times. The different approaches to object tracking comprise pipelines that include various filters which are being applied to one or more of the frames $F_i \in F_{input}$ or to outcome of intermediate steps of the pipeline such as $o_i \in O_{seg}$. Those filters can be of various type such as image processing filters or machine learning algorithms. Let $g_i \in$ V be an indexed pixel in the grid of a frame f consisting of n pixels. The computation of the mean value μ of all pixels in V can be defined as

$$\mu = \frac{1}{n}\sum_{i=1}^{n} \varphi_f(g_i) \tag{3.1}$$

3.6 TENTATIVE TRACK UPDATE

The tentative tracks are created when the detection confidence associated to the tracks is not high enough to confirm the track as a true track. The states of tentative tracks could either be confirmed as a confirmed track or be deleted as a dead track. Observing that if a target has been detected multiple times with low confidence, it could be an actual target. Based on this, a threshold count based confirmation of tentative track is utilized. If the times of a tentative track has been associated with low confidence detections are above the threshold, the track will be updated to be a confirmed track. Tentative tracks will remain their states until this threshold is satisfied. A proper threshold would balance between resisting to the false alarms and reducing missing tracks.

Another threshold for the number of no detection associated to the track is applied for deleting a tentative track. If the unassociated number is beyond the threshold, the track will be deleted, assuming the previous detections associated come from false alarms. This is useful to reduce the computational burden and maintain only few eligible tentative tracks that could be an actual target trajectory. The snakes represent a general technique to match a deformable model to a given frame by minimizing its energy function. The original snake energy term E_{snake} of a contour v(s) as shown in Equation represents a controlled continuity spline under the influence of internal and external forces.

$$E_{snake}(v(s)) = \int_0^1 [E_{int}(v(s)) + E_{img}(v(s)) + E_{con}(v(s))] ds \tag{3.2}$$

Thereby, E_{int} represents the internal energy of the snake, consisting of two weighted terms that make the snake behave like a membrane and like a thin plate respectively. The image forces E_{img} ensure that the snake moves towards line, edge or contour features of the image. The external constraint forces E_{con} eventually locate the snake near the desired local mini- mum.

3.7 TRACK DELETION

Track deletion removes the tracks that are no longer receiving updates from the measurements (i.e. no measurement associated with them) or being out of field of view. Main reason of the track deletion is long term occlusion or target moved out of the scenario. In some applications that require re-identify objects if it disappeared for certain time, such as video surveillance, track deletion should be paid more attention and adapted to these special needs. In the purpose of developing tracking algorithms on ADAS or ADS platforms, maintaining a long history of identities is neither needed, nor computationally efficient.

Table 3.2: Features used to classify track deletion.

Feature	Description
Appearance	Mean normalized correlation coefficients (NCC)

	between associated detection and visual tracking output
Discrimination	Mean NCC between visual tracking output and Mneighbor samples with bounding boxoverlap <0.1.
Boundary	Normalized distance to the closest boundaries
Detection Overlap	Bounding box overlap between the associated de tection and track prediction Confidence
Detection Confidence	of the associated detection
Tracking Confidence	Visual tracking confidence of last frame
Detection Rate	Percentage of detections in last N frame window
Missing Detection Rate	Percentage of detections in last N frame window
Mean Detection Confidence	Mean of detection confidence in last N frame window

4. RESULTS

In this research work an attempt is made for object segmentation and tracking in digital image with active contour model and heuristic method for ground target detection, enough labeled ground truths of objects in the images are needed. In our thesis the VTB-100 datasets are chosen for the experiment. The VTB-100 datasets contain several different challenging benchmark data for multiple vision tasks, like stereo matching, object detection, scene segmentation, object tracking, as well as road/lane detection. The object detection dataset is used for training the proposed object detector and the object tracking dataset is adopted for evaluating and benchmarking the proposed tracking algorithm. The CNN detector is constructed and trained based on one of the popular deep learning framework MXNet developed by the Distributed (Deep) Machine Learning Community (DMLC). The parameters of the network were initialized using a pre-trained ResNet-101 model1 trained on ImageNet for image classification purposes. The model was trained for 10 epochs, that took approximately 8 hours on the aforementioned machine using two GPUs.

The detector generalized well on both VTB-100 testing datasets and unseen datasets for the trained target types. Using VTB-100 detection evaluation tool, the mean average precision (mAP) for 'Car' category by the proposed detector is 86.42%, 89.27%, 72.69% for 'Moderate', 'Easy', 'Hard', while we used active contour model and heuristic method along with Faster R-CNN is 82.11%, 88.70%, 71.19% respectively, showing the effectiveness of the improvements to the original detection algorithm by the proposed detector. However, the detector still has problems with missing some objects and false alarms as well, which remains to be solved by the proposed tracker.

Table 4.1: KITTI benchmark results.

Algorithms	MOTA (%)↑	MOTP (%)↓	FM↓	IDS↓	MT (%)↑	ML (%)↓
MDP	76.6	82.1	387	130	52.1	13.0
SCEA	75.6	79.4	448	104	53.1	11.5
CIWT	75.4	79.3	660	165	49.8	10.3
Ours	65.9	76.7	429	89	50.2	10.8

The evaluation is carried out on the 'car' category. Results are presented in Table 4.1. The proposed algorithm performs good in terms of FM, IDS, MT, and ML, which indicates the effectiveness of the track management. However, the MOTA and MOTP are not as good as other state-of-art algorithms. Improving the precision and accuracy is one of the direction of future research.

The overall processing time of the proposed algorithm including detection and tracking is 11 FPS on average, however, comparing this with other tracking algorithms is hard due to different implementation and executing machine, and the proposed algorithm parallelized the detection and tracking. Therefore, there is no easy way to only compare the tracking processing time with other algorithms.

4.1 ATTRIBUTE-BASED BENCHMARKING

To address the capability of the proposed tracker and better analyze the performance in terms of advantages and disadvantages, benchmarking is carried out by groups of video sequences classified based on different attributes, as well as an overall averaged performance comparison. The exact grouping of the video sequences to different attributes can be found in The tracker is initialized in the first frame using the ground truth position and run sequentially frame by frame without extra input except for the images after that.

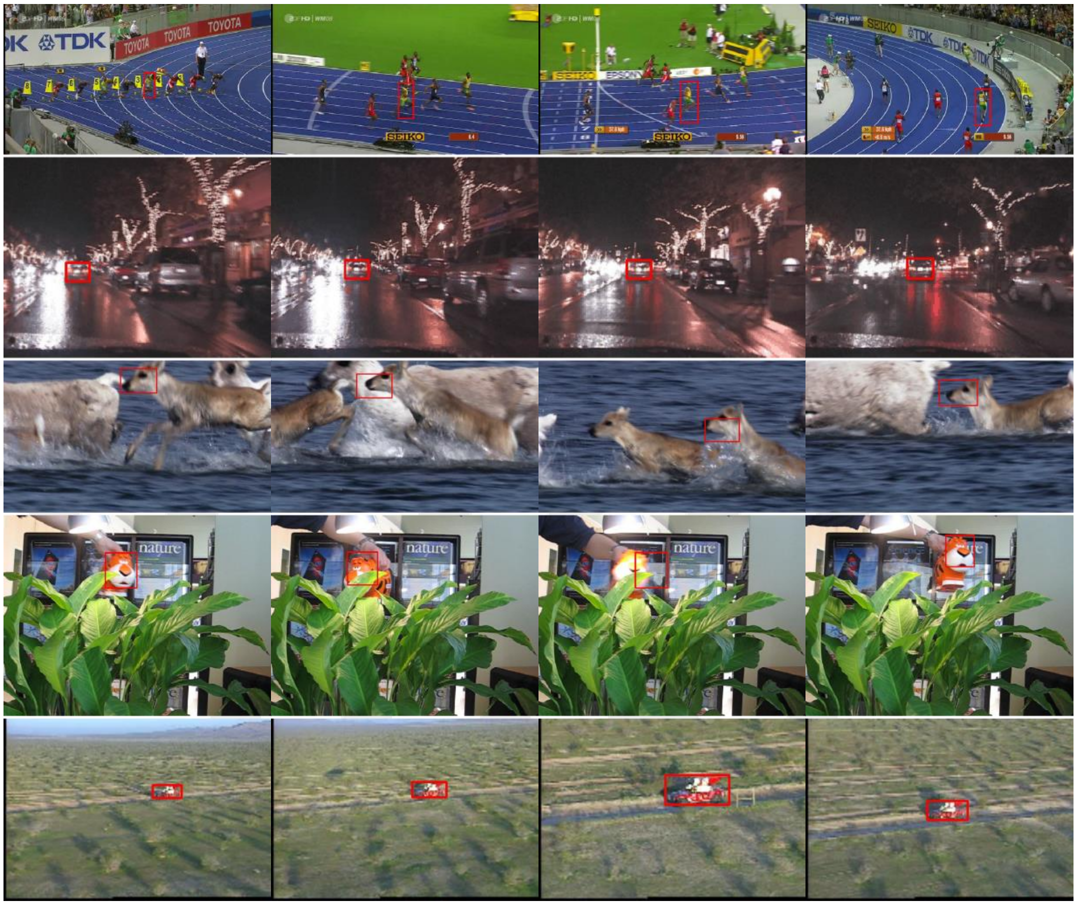


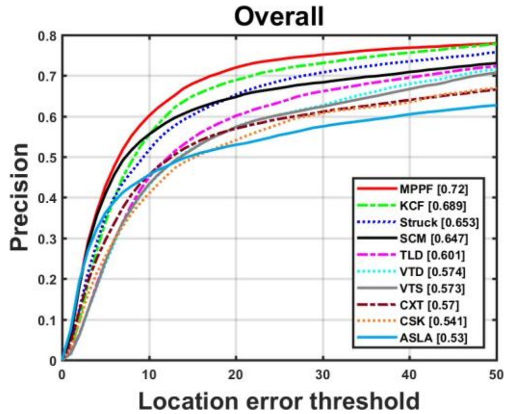
Figure 4.1: Object tracking results (in red rectangle) on VTB-100 datasets (better view in color).

Five typical scenarios in the benchmark is presented in Figure 4.1. The ‘bolt’ data shows changes in object’s own appearance change, from different viewpoint, the proposed tracker adapts to the changes of the objects. In the second sequence ‘Car Dark’, under complex illumination conditions in the dark, the tracker follows the vehicle correctly. In third ‘deer’ sequence, abrupt motion happens throughout the data, and the mixed-motion proposal helps the tracker correctly track the deer head. In the fourth ‘Tiger1’ data, the toy tiger is occluded frequently by the leaves. The occlusion and illumination change caused the tracker drift, however, the tracker successfully recovers from the drift. The last example shows the tracker adaptively follows the scale change of the object. Using the VTB-100 benchmark suite, the proposed algorithm is evaluated quantitatively with recent published algorithms’ results on the benchmark, namely, ASLA, cpf, CSK, CXT, DFT, KCF, LOT, LSK, MIL,MTT, OAB, sbt, Struck, SCM, TLD, TM, VTD and VTS . The proposed tracker name is abbreviated as MMPF.

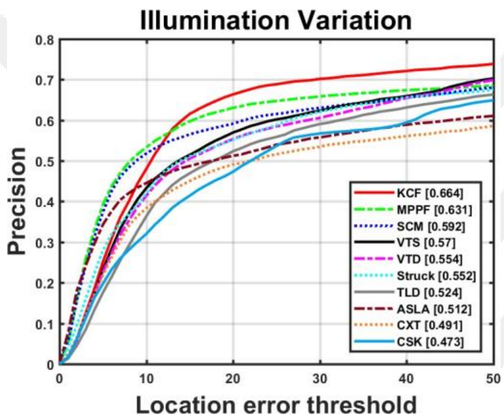
Figure 4.2: Object Tracking results

Figure 4.2 shows the Sample tracking results on dataset, targets are marked with colored rectangles and with numbered ID (better view in color). The image has dimension of 875× 775

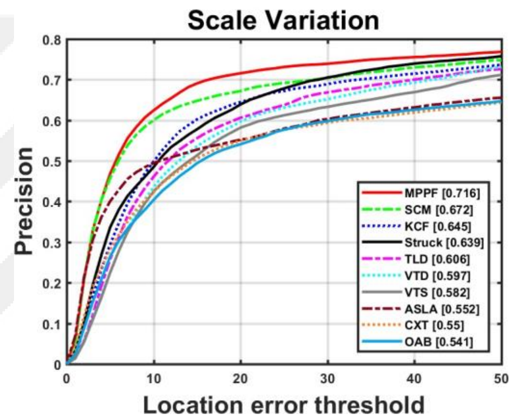
And width of 857 pixels and Height of 775 pixels and bit depth is 32.



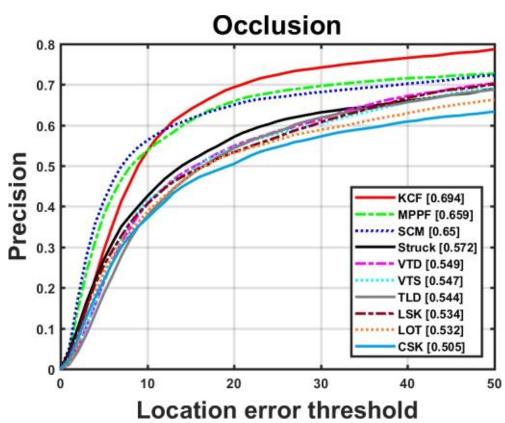
(a) Overall



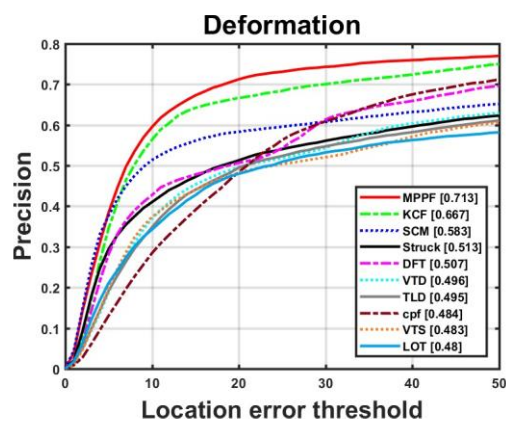
(b) Illumination Variation



(c) Scale Variation



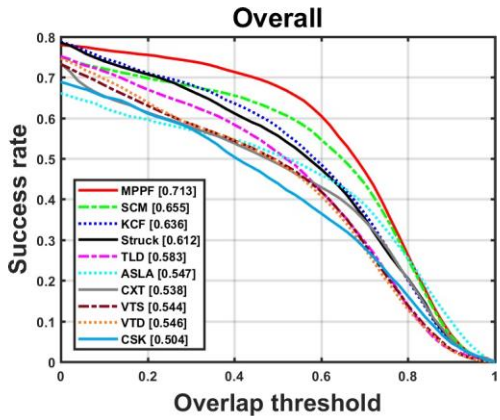
(d) Occlusion



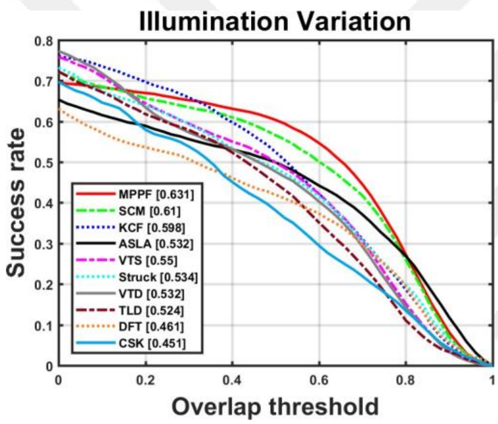
(e) Deformation

Figure 4.3: VTB-100 Tracking Results - Precision plots (better view in color).

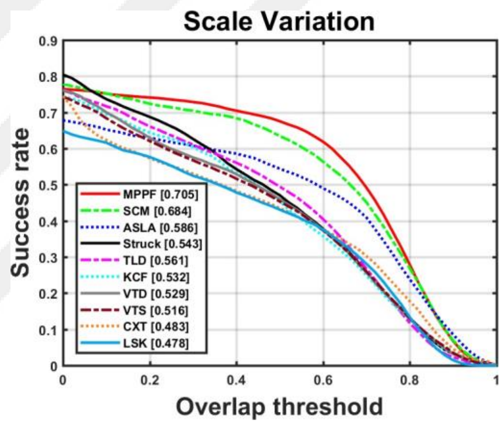
The succession plot is shown in Figure 4.4 and Figure 4.5.



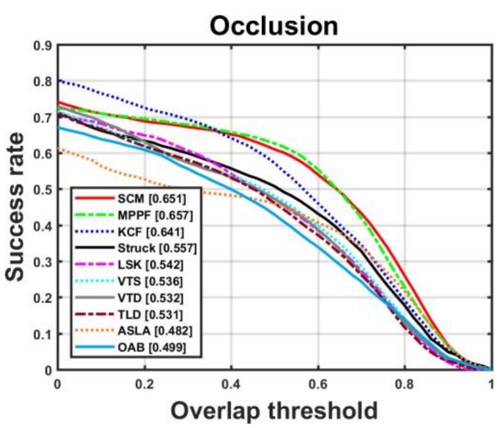
(a) Overall



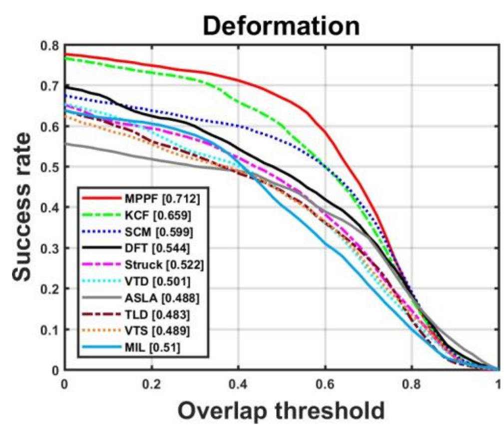
(b) Illumination Variation



(c) Scale Variation

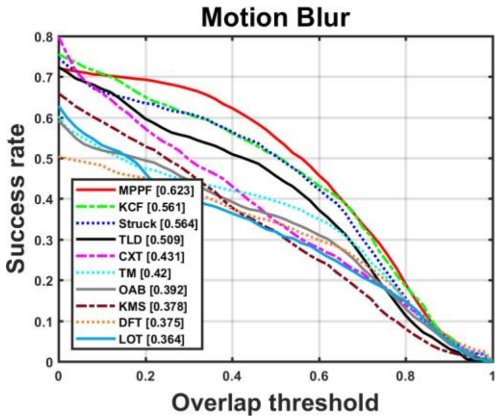


(d) Occlusion

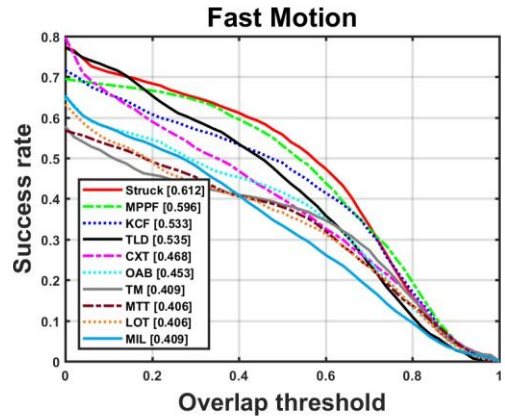


(e) Deformation

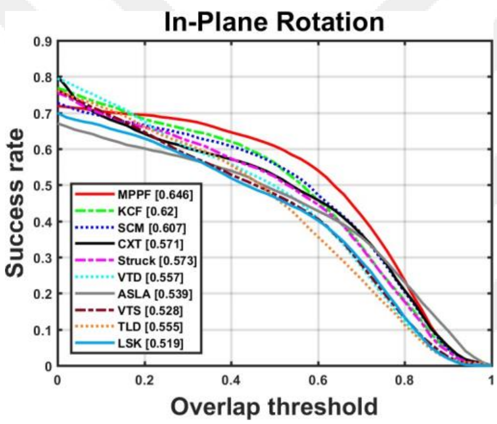
Figure 4.4: VTB-100 Tracking Results - Success plots (better view in color).



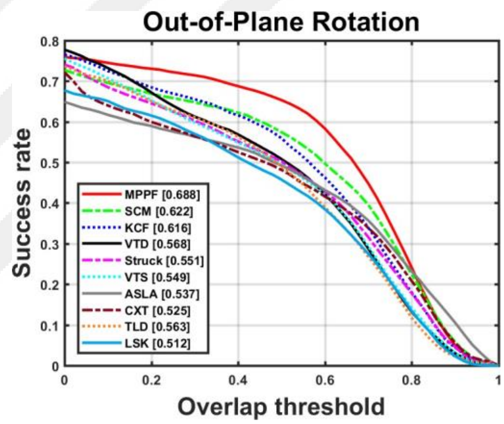
(a) Motion Blur



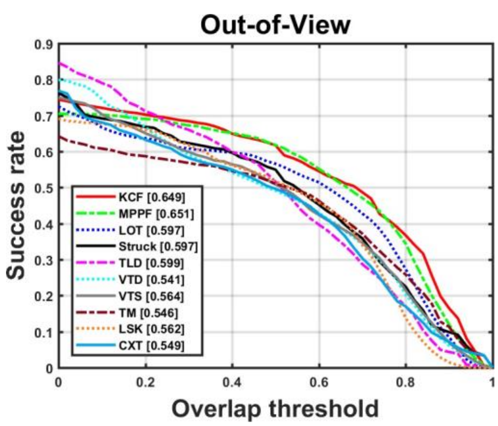
(b) Fast Motion



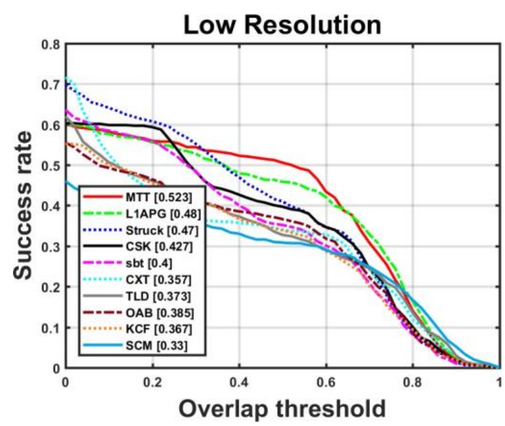
(c) In-Plane Rotation



(d) Out-of-Plane Rotation



(e) Out-of-View



(f) Low Resolution

Figure 4.5: VTB-100 Tracking Results - Success plots (better view in color) cont.

Overall, the proposed algorithm performs best in terms of center error and bounding box overlap as the ranks shown in the precision and success plots. In figure 4.5 the images has dimension of 763×638 and width of 763 pixels and Height of 638 pixels and bit depth is 24.

The proposed algorithm achieves highest rank in 6 different sub-categories and 3 second ranks for fast motion in the success plots, and the algorithm performs poorly in the low resolution sequences. However, comparing to other categories, all trackers suffer from low resolution and perform worse than in other categories. The reason behind this is because less information can be extracted to localize the targets when the image resolution is low. This quantitative benchmarking shows the good accuracy of the proposed tracker comparing to the state-of-art published algorithms. The proposed tracker is designed to achieve high speed without losing the accuracy. The speed of the algorithm is compared with the classical KCF algorithm from the popular correlation filter family on the test machine. The proposed algorithm can achieve an average of 400 FPS while KCF algorithm can only perform 41 FPS on average, showing the proposed tracker is almost 10x faster than the popular KCF algorithm. The speed is crucial to apply the algorithm in multiple target tracking scenarios presented in the following chapter, which will increase the processing time linearly to the number of targets.

NfS Datasets Since high frame rate is achievable by the proposed algorithm, a high frame rate video dataset called NfS is being also tested with. Each sequence in the NfS datasets contains a low and a high frame rate video with 30 FPS and 240 FPS respectively. The 240 FPS sequences are utilized for benchmarking with the proposed algorithm to test the high frame rate performance, and according to the processing time from previous tests, with such a high frame rate video, the algorithm should still achieve real-time processing throughput with no delay.

Result Analysis The tracking algorithm is compared with algorithms, namely KCF, Staple and CFwLB, that processing more than 50 FPS according to the claimed benchmark results provided in. The precision and success plots are shown in Figure 4.6 and Figure 4.7 respectively.

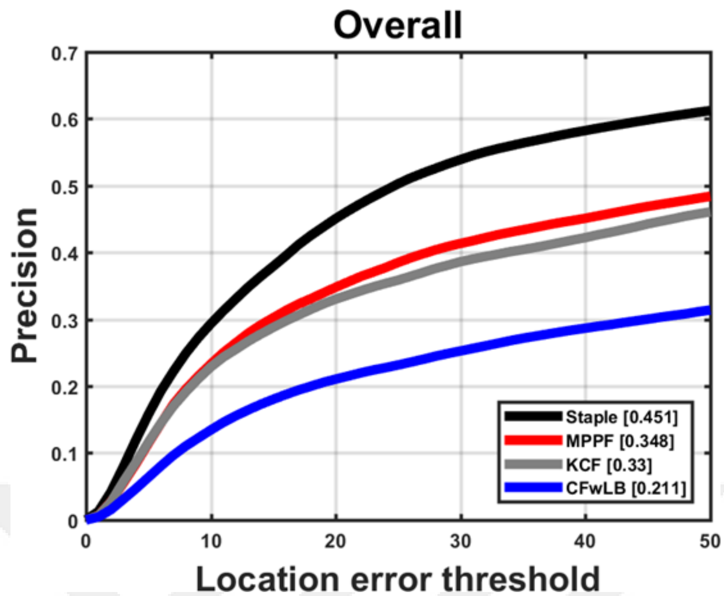


Figure 4.6: NfS Tracking Results - Precision plots (better view in color).

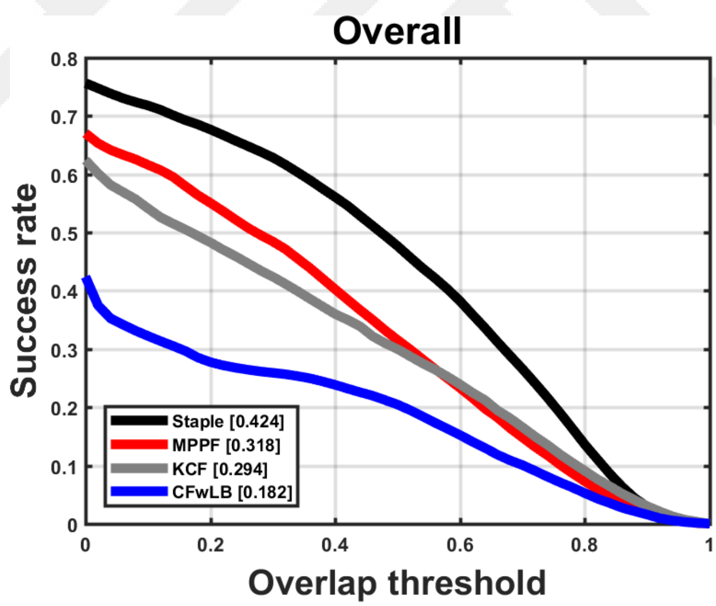


Figure 4.7: NfS Tracking Results - Success plots (better view in color).

According to the speed for the given algorithms are claimed as KCF (170.4 FPS), CFwLB (83 FPS) and Staple (50.8 FPS) in their tests. NfS Tracking Results image has dimension of 554× 454 And width of 554 pixels and Height of 454 pixels and bit depth is 32.

Still, the proposed tracker (400 FPS) outperforms the mentioned algorithms in speed with a large margin while maintaining the comparable accuracy. A high-speed visual target tracker is presented in this chapter, without losing the performance in terms of accuracy. The proposed tracker utilizes a mixed-motion proposal particle filter, and employs the Ridge Regression to learn and update a robust appearance model. Occlusion and partial out-of-scene is also handled in the tracker. Experiments and comparison on public dataset shows the proposed tracker achieves best performance and outperforms other algorithms in speed with a large margin.

5. CONCLUSION

In this thesis, object segmentation and tracking in digital image with Active Contour model and heuristic method and along with it multiple ground target tracking algorithms are studied. Depending on the platforms, three different types of tracking algorithms are proposed according to the specialties of these platforms. The dependent target tracking for ground targets is studied in in this thesis and Traditional target tracking algorithms assume a state independence among targets, this will result in erroneous tracks for ground target that have dependent motion due to scenario constraints. The proposed algorithm models the states of the target jointly and derives the MRF based probabilistic data association filter. The state dependency is modeled using the driving behavior models that captures the specialty of ground target motion pattern and dependency among the targets. The experiments and simulations show the effectiveness of such a modeling that improves the predictions, and the proposed behavior model combined with MRF-PDA can reduce the false associations and improve performance evaluated using RMSE metric compared with classical JPDA.

To track ground targets in the urban scene, camera is an effective and low-cost sensor choice that can record high detailed information about the targets. To build a multiple ground target tracking algorithm using a single camera, the single target tracking is firstly studied. To achieve real-time processing on multiple targets, the speed of single target tracking is one of the significant part. With speed requirement in mind, the proposed algorithm adopts particle filtering with resizing templates to replace the commonly used exhaustive searching in different scales to localize the potential target candidates. To further increase the effectiveness of the limited number of particles, a mix-state proposal distribution is adopted. The robust appearance model is based on Ridge Regression that learns to discriminate the target and background. The appearance model is fast to train and easily update according to the track confidence. Extensive experiments are carried out with public benchmarking datasets. The results of the experiment show the proposed targets achieves good performance while beats other algorithms in speed with a large margin.

Despite the single target tracking algorithm developed, the multiple target tracking algorithm still needs to solve the detection generation, data association and track management problems. A new tracking algorithm dealing with these problems are addressed. Recent advances of CNN have drastically improved the real-time object detection tasks in images. The proposed algorithm combines the Res-Net with RPN to train an object detector and generate the detections of interested ground targets. The detector is further improved by using perspective information in the urban scene. Since the detector generates not only the detections, but also the class information, a new multi-class data association algorithm using a hierarchical object class tree is proposed. The high accuracy CNN detectors help improve track management by using the detection confidence and a new SVM based track deletion is also proposed to correctly remove the dead tracks.

5.1 FUTURE RESEARCH DIRECTION AND DISCUSSION

Some future work is born from the current research topics, and there are still space to improve the presented work to deploy them into real world systems. Also adapting to a real-world system has many requirements and limitations that is different from the experimental setups, such as processing time limits, computational resource constraints, and even power consumption restrictions. Several improvements could be made to expand the proposed research:

5.1.1 Fusing with More Sensors

In all three works, only single sensor model was considered. To improve the accuracy, stability, coverage, and robustness, multiple sensors could be employed. Potential sensors to be fused could be cameras, stereo cameras, mid-range radars and LiD-ARs. Comparing to using single sensor for vehicle tracking, fusion of multiple sensors would increase the performance by combining different source of information. Fusion algorithms introduces new problems to be handled, such as sensor registration, sensor bias estimation, and fusion algorithm design. These problems need to be solved in order to achieve the benefits of fusing multiple sensors. Optimizing for Embedded Systems. In real world applications, such automotive, video surveillance and drone navigation, many platforms are implemented on embedded systems. Such systems usually have limited resources for computing and restricted power supply. Comparing to a regular PC machine, the CPU/RAM/GPU of embedded systems are far less powerful in terms of clock speed, speed-up intrinsic, cache/memory size, bus speed and computational cores, and the power consumption may also be limited. These resource limits would slow the proposed algorithms down and eventually reduce the speed as well as accuracy, making the algorithms not deployable. Trade-offs needs to be made for optimizing the usage of the proposed algorithms in embedded systems. Potential improvements would be developing special hard- ware speedup units for some modules, such as CNN based detection algorithms. In the algorithm side, designing new detection methods would significantly reduce the computational load.

5.1.2 Application of Generated Tracks

Application of generated tracks is the next steps of ground target tracking algorithm. Tracks themselves contains useful information like speed and location that can be utilized immediately, however, more useful information can be extracted based on the tracks. And applications of these tracks for other tasks could be studied. ADAS and ADS In the automotive area, the output of extracted tracks can be utilized to help implement forward collision warning, perform adaptive cruise control, and avoid crash in ADAS and ADS. Intelligent Transportation System and Traffic Control Using traffic surveillance cameras or large area surveillance platforms on helicopters or drones, the pro- posed algorithms could be applied to automatically generate traffic reports and jam alerts, traffic predictions, and intelligent traffic controls. Robotics and Its Navigation Vision based navigation in both indoor and outdoor scenarios requires correct tracking and prediction of surrounding obstacle targets. The generated tracks from the camera mounted on robots helps path planning, obstacle avoidance and navigation.

REFERENCES

[1] T. Mendonça, P.M. Ferreira, J.S. Marques, A.R. Marcal, J. Rozeira, “dermoscopic image database for research and benchmarking“. In Proceedings of the 2013 35th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), Osaka, Japan, 3–7 July 2013; pp. 5437–5440.

[2] E Meijering, I Smal, G. Danuser “Tracking in molecular bioimaging“. IEEE Signal Proc Mag. 2016; 23:46–53.

[3] B.H. Menze, A. Jakab, S. Bauer, J. Kalpathy-Cramer, K. Farahani, J. Kirby, Y.Burren, N. Porz, J. Slotboom, R. Wiest, “The multimodal brain tumor image segmentation benchmark (BRATS) “. IEEE Trans. Med. Imaging 2014, 34, 1993–2024.

[4] C. Li, C. Xu, M.D. Gui Fox, “Distance regularized level set evolution and its application to image segmentation“. IEEE Trans. Image Process. 2010, 19, 3243–3254

[5] X. Hou, L. Zhang, Saliency detection: “A spectral residual approach“. In Proceedings of the 2007 IEEE Conference on Computer Vision and Pattern Recognition, Minneapolis, MN, USA, 17–22 June 2007; pp. 1–8.

[6] T.F. Chan, L.A. Vese, “Active contours without edges“. IEEE Trans. Image Process. 2001, 10, 266–277.

[7] C. Li, C.-Y. Kao, J.C. Gore, Z. Ding, “Implicit Active Contours Driven by Local Binary Fitting Energy“ In Proceedings of the 2007 IEEE Conference on Computer Vision and Pattern Recognition, Minneapolis, MN, USA, 17–22 June 2007; Institute of Electrical and Electronics Engineers (IEEE): Piscataway, NJ, USA, 2007; pp. 1–7

[8] A. Tsai, A. Yezzi, A.S. Willsky, “Curve evolution implementation of the Mumford-Shah functional for image segmentation, denoising, interpolation, and magnification“. IEEE Trans. Image Process. 2001, 10, 1169–1186.

[9] CJ Veenman, MJT Reinders, E. Backer “Resolving Motion Correspondence for Densely Moving Points“. IEEE T Pattern Anal. 2016; 23:54–72.

[10] L. Itti, C. Koch, E. Niebur, “A model of saliency-based visual attention for rapid scene analysis“. IEEE Trans. Pattern Anal. Mach. Intell. 1998, 20, 1254–1259

[11] M.-M. Cheng, N.J. Mitra, X. Huang, P.H. Torr, S.-M. Hu, “Global contrast based salient region detection“. IEEE Trans. Pattern Anal. Mach. Intell. 2015, 37, 569–582.

[12] J. Lie, M. Lysaker, X.-C. Tai, “A binary level set model and some applications to Mumford-Shah image segmentation“. IEEE Trans. Image Process. 2006, 15, 1171–1181.

[13] G. Carneiro, B. Georgescu, S. Good, and D. Comaniciu, “Detection and Measurement of Fetal Anatomies from Ultrasound Images using a Constrained Probabilistic Boosting Tree, Medical Imaging“, IEEE Transactions on, vol. 27, no. 9, pp. 1342–1355, 2018.

[14] B. Wang, X. Gao, D. Tao, X. Li, “A nonlinear adaptive level set for image segmentation“. IEEE Trans. Cybern. 2014, 44, 418–428.

[15] M. C. Chi, C. H. H. Yeh, and M. J. J. Chen, “Robust Region-of-Interest Determination Based on User Attention Model Through Visual Rhythm Analysis,” Circuits and Systems for Video Technology, IEEE Transactions on, vol. 19, no. 7, pp. 1025–1038, 2019.

[17] D. Comaniciu, V. Ramesh, and P. Meer, “Kernel-based object tracking,” IEEE Transac- tions on Pattern Analysis and Machine Intelligence, vol. 25, no. 5, pp. 564–577, 2003.

[18] D. Comaniciu, P. Meer, “Mean shift analysis and applications,” in Computer Vision, 1999. The Proceedings of the Seventh IEEE International Conference on, vol. 2, 2016, pp. 1197–1203.

[19] L. Weruaga, R. Verdu, J. Morales, “Frequency domain formulation of active parametric deformable models“. IEEE Trans. Pattern Anal. Mach. Intell. 2004, 26, 1568–1578

[20] Y. Qi, S. Zhang, F. Jiang,; H. Zhou,; D. Tao,; X. Li, “Siamese local and global networks for robust face tracking“. IEEE Trans. Image Process. 2020, 29, 9152–9164.

[21] M. N. Do, M. Vetterli, “The contourlet transform: an efficient directional multireso- lution image representation” IEEE transactions on image processing : a publication of the IEEE Signal Processing Society, vol. 14, no. 12, pp. 2091–2106, 2015.

[22] L. Zhang, S. Zhang, F. Jiang, Y. Qi, J. Zhang,; Guo, Y. Zhou, H. BoMW “Bag of manifold words for one-shot learning gesture recognition from kinect“. IEEE Trans. Circuits Syst. Video Technol. 2017, 28, 2562–2573.

[23] Y. Qi, H. Yao, X. Sun, X. Sun, Y. Zhang, Q. Huang, “Structure-aware multi-object discovery for weakly supervised tracking“. In Proceedings of the 2014 IEEE International Conference on Image Processing (ICIP), Paris, France, 27–30 October 2014; pp. 466–470.

[24] M. Elad and M. Aharon, “Image denoising via sparse and redundant representations over learned dictionaries,” IEEE transactions on image processing : a publication of the IEEE Signal Processing Society, vol. 15, no. 12, pp. 3736–3745, 2016.

[25] D. Exner, E. Bruns, D. Kurz, Grundh&#x00F6, A. fer, and O. Bimber, “Fast and robust CAMShift tracking,” in Computer Vision and Pattern Recognition Workshops (CVPRW), 2010 IEEE Computer Society Conference on, 2010, pp. 9–16.

[26] K. Fukunaga, L. D. Hostetler, “The estimation of the gradient of a density function, with applications in pattern recognition,” Information Theory, IEEE Transactions on, vol. 21, no. 1, pp. 32–40, 1975.

[27] G.R. Reddy, K. Ramudu, P. Yugander, R.R. Rao, “Fast global region based minimization of satellite and medical imagery with geometric active contour and level set evolution on noisy images“. In Proceedings of the 2011 IEEE Recent Advances in Intelligent Computational Systems, Trivandrum, India, 22–24 September 2011; pp. 696–700.

[28] Cohen, L.D.; Cohen, I. “Finite-element methods for active contour models and balloons for 2-D and 3-D images“. IEEE Trans. Pattern Anal. Mach. Intell. 1993, 15, 1131–1147.

[29] G. Gualdi, A. Prati, and R. Cucchiara, “Multistage Particle Windows for Fast and Accurate Object Detection,” IEEE Transactions on Pattern Analysis and Machine In- telligence, vol. 34, no. 8, pp. 1589–1604, 2012.

[30] S. Gunn, M. Nixon, “A robust snake implementation; a dual active contour,” IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 19, no. 1, pp. 63–68, 2016.

[31] F. Perazzi, P. Krähenbühl, Y. Pritch, A. Hornung, "Saliency filters Contrast based filtering for salient region detection". In Proceedings of the 2012 IEEE Conference on Computer Vision and Pattern Recognition, Providence, RI, USA, 16–21 June 2012; Institute of Electrical and Electronics Engineers (IEEE): Piscataway, NJ, USA, 2012; pp. 733–740

[32] A. Kolarow, M. Brauckmann, M. Eisenbach, K. Schenk, E. Einhorn, K. Debes, and H. M. Gross, "Vision-based hyper-real-time object tracker for robotic applications," in Intelligent Robots and Systems (IROS), 2012 IEEE/RSJ International Conference on, 2012, pp. 2108–2115.

[33] R. Achanta, S. Susstrunk, "Saliency detection using maximum symmetric surround". In Proceedings of the 2010 IEEE International Conference on Image Processing, Hong Kong, China, 26–29 September 2010; pp. 2653–2656.

[34] S. Goferman, L. Zelnik-Manor, A. Tal, "Context-aware saliency detection". IEEE Trans. Pattern Anal. Mach. Intell. 2011, 34, 1915–1926

[35] Ç. Aytekin, E.C. Ozan, S. Kiranyaz, M. Gabbouj, "Visual saliency by extended quantum cuts". In Proceedings of the 2015 IEEE International Conference on Image Processing (ICIP), Quebec City, QC, Canada, 27–30 September 2015; pp. 1692–1696.

[36] E. Le Pennec, S. Mallat, "Sparse geometric image representations with bandelets," IEEE transactions on image processing : a publication of the IEEE Signal Processing Society, vol. 14, no. 4, pp. 423–438, 2015.

[37] K. Fu, C. Gong, I.Y.-H. Gu, J. Yang, "Normalized cut-based saliency detection by adaptive multi-level region merging". IEEE Trans. Image Process. 2015, 24, 5671–5683

[38] D. Martin, C. Fowlkes, D. Tal, J. Malik, "A Database of Human Segmented Natural Images and Its Application to Evaluating Segmentation Algorithms and Measuring Ecological Statistics". In Proceedings of the Eighth IEEE International Conference on Computer Vision, Vancouver, BC, Canada, 7–14 July 2001.

[39] J. J. Shi, Q. Yan, L. Xu, J. Jia, "Hierarchical image saliency detection on extended cssd". IEEE Trans. Pattern Anal. Mach. Intell. 2015, 38, 717–729

[40] C. Li, C.-Y. Kao, J.C. Gore, Z. Ding, “Minimization of region-scalable fitting energy for image segmentation“. IEEE Trans. Image Process. 2008, 17, 1940–1949

[41] J. Mairal, M. Elad, and G. Sapiro, “Sparse representation for color image restoration,” IEEE transactions on image processing : a publication of the IEEE Signal Processing Society, vol. 17, no. 1, pp. 53–69, 2018.

[42] S. Lankton, A. Tannenbaum, “Localizing region-based active contours“. IEEE Trans. Image Process. 2008, 17, 2029–2039.

[43] G. Gilboa, N. Sochen, Y. Zeevi, “Image enhancement and denoising by complex diffusion processes“. IEEE Trans. Pattern Anal. Mach. Intell. 2004, 26, 1020–1036

[44] X. Mei, H. Ling, “Robust Visual Tracking and Vehicle Classification via Sparse Representation,” IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 33, no. 11, pp. 2259–2272, 2011.