

Institutionen för systemteknik

Department of Electrical Engineering

Examensarbete

Geometric Scene Labeling for Long-Range Obstacle Detection

Examensarbete utfört i Datateknik
vid Tekniska högskolan vid Linköpings universitet
av

Patrik Hillgren

LiTH-ISY-EX--14/4819--SE

Linköping 2014



Linköpings universitet
TEKNISKA HÖGSKOLAN

Geometric Scene Labeling for Long-Range Obstacle Detection

Examensarbete utfört i Datateknik
vid Tekniska högskolan vid Linköpings universitet
av


Patrik Hillgren

LiTH-ISY-EX--14/4819--SE

Handledare: **Peter Pinggera**
Daimler AG
Rudolf Mester
ISY, Linköpings universitet

Examinator: **Michael Felsberg**
ISY, Linköpings universitet

Linköping, 9 september 2014

	Avdelning, Institution Division, Department Computer Vision Laboratory Department of Electrical Engineering SE-581 83 Linköping	Datum Date 2014-09-09
---	--	--

Språk Language <input type="checkbox"/> Svenska/Swedish <input checked="" type="checkbox"/> Engelska/English <input type="checkbox"/> _____	Rapporttyp Report category <input type="checkbox"/> Licentiatavhandling <input checked="" type="checkbox"/> Examensarbete <input type="checkbox"/> C-uppsats <input type="checkbox"/> D-uppsats <input type="checkbox"/> Övrig rapport <input type="checkbox"/> _____	ISBN _____ ISRN LiTH-ISY-EX--14/4819--SE Serietitel och serienummer ISSN Title of series, numbering _____
URL för elektronisk version http://urn.kb.se/resolve?urn=urn:nbn:se:liu:diva-XXXXX		

Titel Title Författare Author	Geometrisk Segmentering för Detektion av Objekt på Stora Avstånd Geometric Scene Labeling for Long-Range Obstacle Detection Patrik Hillgren
--	---

Sammanfattning Abstract <p>Autonomous Driving or self driving vehicles are concepts of vehicles knowing their environment and making driving manoeuvres without instructions from a driver. The concepts have been around for decades but has improved significantly in the last years since research in this area has made significant progress. Benefits of autonomous driving include the possibility to decrease the number of accidents in traffic and thereby saving lives.</p> <p>A major challenge in autonomous driving is to acquire 3D information and relations between all objects in surrounding traffic. This is referred to as <i>spatial perception</i>. Stereo camera systems have become a central sensor module for advanced driver assistance systems and autonomous driving. For object detection and measurements at large distances stereo vision encounter difficulties. This includes objects being small, having low contrast and the presence of image noise. Having an accurate perception of the environment at large distances is however of high interest for many applications, especially autonomous driving.</p> <p>This thesis proposes a method which tries to increase the range to where generic objects are first detected using a given stereo camera setup. Objects are represented by planes in 3D space. The input image is segmented into the various objects and the 3D plane parameters are estimated jointly. The 3D plane parameters are estimated directly from the stereo image pairs. In particular, this thesis investigates methods to introduce geometric constraints to the segmentation or labeling task, i.e assigning each considered pixel in the image to a plane.</p> <p>The methods provided in this thesis show that despite the difficulties at large distances it is possible to exploit planar primitives in 3D space for obstacle detection at distances where other methods fail.</p>

Nyckelord Keywords	Computer Vision, Autonomous Driving, Stereo Vision, Graph Cuts, Stixels
------------------------------	---

Abstract

Autonomous Driving or self driving vehicles are concepts of vehicles knowing their environment and making driving manoeuvres without instructions from a driver. The concepts have been around for decades but has improved significantly in the last years since research in this area has made significant progress. Benefits of autonomous driving include the possibility to decrease the number of accidents in traffic and thereby saving lives.

A major challenge in autonomous driving is to acquire 3D information and relations between all objects in surrounding traffic. This is referred to as *spatial perception*. Stereo camera systems have become a central sensor module for advanced driver assistance systems and autonomous driving. For object detection and measurements at large distances stereo vision encounter difficulties. This includes objects being small, having low contrast and the presence of image noise. Having an accurate perception of the environment at large distances is however of high interest for many applications, especially autonomous driving.

This thesis proposes a method which tries to increase the range to where generic objects are first detected using a given stereo camera setup. Objects are represented by planes in 3D space. The input image is segmented into the various objects and the 3D plane parameters are estimated jointly. The 3D plane parameters are estimated directly from the stereo image pairs. In particular, this thesis investigates methods to introduce geometric constraints to the segmentation or labeling task, i.e assigning each considered pixel in the image to a plane.

The methods provided in this thesis show that despite the difficulties at large distances it is possible to exploit planar primitives in 3D space for obstacle detection at distances where other methods fail.

Sammanfattning

En autonom bil innebär att bilen har en uppfattning om sin omgivning och kan utifrån det ta beslut angående hur bilen ska manövreras. Konceptet med självkörande bilar har existerat i årtionden men har utvecklats snabbt senaste åren sedan billigare datorkraft finns lättare tillgänglig. Fördelar med autonoma bilar innebär bland annat att antalet olyckor i trafiken minskas och därmed liv räddas.

En av de största utmaningarna med autonoma bilar är att få 3D information och relationer mellan objekt som finns i den omgivande trafikmiljön. Detta kallas för spatial perception och innebär att detektera alla objekt och tilldela en korrekt position till dem. Stereo kamerasystem har fått en central roll för avancerade förarsystem och autonoma bilar. För detektion av objekt på stora avstånd träffar stereo system på svårigheter. Detta inkluderar väldigt små objekt, låg kontrast och närvaron av brus i bilden. Att ha en ackurativ perception på stora avstånd är dock vitalt för många applikationer, inte minst autonoma bilar.

Den här rapporten föreslår en metod som försöker öka avståndet till där objekt först upptäcks. Objekt representeras av plan i 3D rummet. Bilder givna från stereo par segmenteras i olika objekt och plan parametrar estimeras samtidigt. Planens parametrar estimeras direkt från stereo bild paren. Den här rapporten utreder metoder att introducera geometriska begränsningar att använda vid segmenteringsuppgiften.

Metoderna som presenteras i denna rapport visar att trots den höga närvaron av brus på stora avstånd är det möjligt att estimerar geometriska objekt som är starka nog att möjliggöra detektion av objekt på ett avstånd där andra metoder misslyckas.

Acknowledgments

I would like to thank all the people contributing to this thesis. This includes the lovely people of the environment perception department at Daimler AG and my supervisor and examiner at Linköping University. A special thanks goes to my supervisor at Daimler AG, Peter Pinggera, who gave excellent guidance and support throughout the thesis.

Linköping, Oktober 2014
Patrik Hillgren

Contents

Notation	xi
1 Introduction	1
1.1 Background	1
1.1.1 Stereo Vision	1
1.1.2 Stixel World	2
1.1.3 Aim	5
1.2 Problem	6
1.3 Delimitations	7
1.4 Contributions	7
1.5 Report Outline	8
2 Theoretical Background	9
2.1 Stereo Vision	9
2.1.1 Local Methods	10
2.1.2 Global Methods	10
2.2 Probabilistic Graphical Models	11
2.2.1 Factor Graphs	11
2.3 Markov Random Fields	12
2.3.1 Conditional Random Fields	13
2.4 Modeling MRFs	13
2.4.1 Potts Model	14
2.4.2 User Defined Models	15
2.5 Inference Algorithms on Pairwise MRFs	15
2.5.1 Graph cuts	15
2.5.2 Submodularity	17
2.6 Limitations of MRFs	18
3 Energy Formulation	19
3.1 Joint Labeling and Parameter Estimation	19
3.1.1 Formulation of the Different Terms	20
3.1.2 Inference in Practice	21

4	Implementation	23
4.1	Modeling Label Priors	23
4.1.1	Applied Cost Function	23
4.1.2	Relations between PSEs	24
4.1.3	Relations between PSEs and estimated horizon	30
4.1.4	Intensity Differences	30
4.2	Initialization of Graph Cuts	31
4.2.1	Rectangular Grid	32
4.2.2	Superpixel Segmentation (gSLIC)	32
4.2.3	Position Dependent Initialization	32
4.3	Restriction of Expansion Areas	33
4.3.1	Expansion to Neighboring Planar Scene Elements	34
4.3.2	Expansion to Rectangular Area	34
4.3.3	Expansion limited by Horizon	34
4.3.4	Expansion limited by ground	35
5	Results	37
5.1	MRF Results	37
5.2	CRF Results	38
5.3	Evaluation	39
5.3.1	Ground Truth Dataset	39
5.3.2	Segmentation Accuracy and Object Detection Rate	39
5.3.3	Freespace Estimation	40
6	Discussion and Future Work	45
6.1	Method	46
6.2	Results	46
6.3	Future Work	46
6.4	Ethical and Societal Aspects	47
6.5	Conclusions	48
A	Labeling with Dynamic Programming	51
A.1	Tiered Scene Labeling	52
A.2	Fast Tiered Labeling	52
A.3	Tiered Scene Labeling Results	54
	Bibliography	55

Notation

SYMBOLS USED

Symbol	Meaning
E	Energy function
D	Data cost/Unary potential
V	Label inconsistency cost/Pair potential
\vec{I}	Set of Input data
$\vec{\mathcal{L}}$	Set of discrete random variables
$\vec{\ell}$	Realization of $\vec{\mathcal{L}}$
$p(\vec{I} \vec{\ell}, \vec{w})$	Data likelihood
$p(\vec{\ell} \vec{w})$	Label prior
$p(\vec{w})$	Parameter prior
\mathcal{N}	Neighborhood
\mathcal{C}	Set of all cliques
c	clique
W	Width of ROI
H	Height of ROI
ω	Potts strength
p	pixel position (vertex)
q	pixel position (vertex)
l_p	Scene element at pixel position (vertex) p
l_q	Scene element at pixel position (vertex) q
Z	Distance in meters
Z_p	Distance to planar scene element l_p
L_{PSE}	Likelihood function given by relations between PSEs
L_{int}	Likelihood function given by intensity values
f	Focal length of the camera
b	Base of the camera
ϵ_d	Disparity error
ϵ_z	Metric distance error
ϵ_Δ	Metric offset value
Δ_o	Minimum distance allowed between objects
o	Planar scene element representing objects
g	Planar scene element representing ground
s	Planar scene element representing sky/background

ABBREVIATIONS

Abbreviation	Meaning
PSE	Planar Scene Element
DP	Dynamic Programming
CRF	Conditional Random Field
FPGA	Field Programmable Gate Array
MRF	Markov Random Field
SGM	Semi-Global Matching
ROI	Region of interest
RANSAC	Random Sample Consensus
IU	Intersection-over-Union
BN	Bayesian Network
SLIC	Simple Linear Iterative Clustering
gSLIC	GPU Simple Linear Iterative Clustering
GBIS	Graph-Based Image Segmentation
MAP	Maximum a Posterior

1

Introduction

1.1 Background

Autonomous driving or self driving vehicles are concepts of a vehicle knowing its environment and making decisions without instructions from a driver. To be able to interpret the environment multiple techniques are used. These include radar and laser measurements, GPS information and computer vision technologies. The concept of autonomous driving has been around for decades but research has made significant progress in the last years [28]. Benefits of autonomous driving include the possibility to decrease the number of accidents in traffic and thereby saving lives. Accidents today are mainly caused by a human factor, people reacting too slowly to the traffic situation or not paying enough attention to the road. One of the goals of autonomous driving is to reduce the human factor in everyday traffic and thereby reducing the numbers of accidents. Autonomous driving is also a very convenient function, making driving in general easier.

1.1.1 Stereo Vision

A major challenge in autonomous driving is to acquire information of and relations between all objects in surrounding traffic. This is referred to as *spatial perception*. Spatial perception of a driving environment can be achieved with varying quality depending on what technique is used. Existing methods use radar and laser measurements and computer vision techniques. A computer vision approach to obtain spatial perception is *Stereo Vision*. Stereo vision is the concept of having two cameras at slightly different positions with a known relative displacement to each other. The displacement of image points between image pairs, called *disparity*, can be calculated and becomes smaller with increasing distance to the scene point. Disparities between points in stereo image pairs can be estimated with various methods and enable triangulation of 3D points (see Figure

1.1). Disparity estimation is an optimization problem and methods differ in the formulation of the objective function and optimization method. Methods which enable disparity measurements at every image point provide *dense stereo matching*. Methods which enable disparity measurements at extracted feature points give *sparse stereo matching*.

Semi-Global Matching

One method to obtain depth information from stereo image pairs is *Semi-Global Matching* (SGM [16]). SGM provides dense stereo matching (see Figure 1.2) by optimizing a combination of matching costs and smoothness constraints in an efficient way. Performing SGM in real-time on a low-powered inexpensive *Field Programmable Gate Array* (FPGA) was first introduced in 2008 [13], introducing new possibilities for future use of autonomous vehicles.

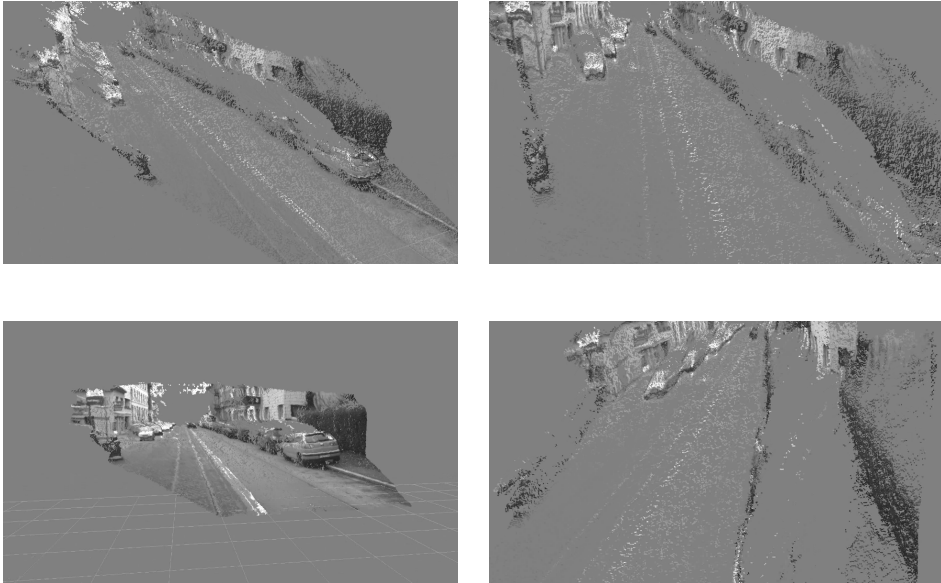


Figure 1.1: Point cloud obtained by triangulating 3D points. The images are showing results from one stereo image pair at different angles using SGM.

1.1.2 Stixel World

Depth information generated by SGM contains a large amount of data which is not optimal for further processing steps. As can be seen in Figure 1.2 almost every pixel is assigned a depth value. In 2009 the *stixel world* [3] was introduced. The stixel world uses depth information obtained from SGM to build a world of rectangular objects, called *stixels*. Stixels are perpendicular to the ground and

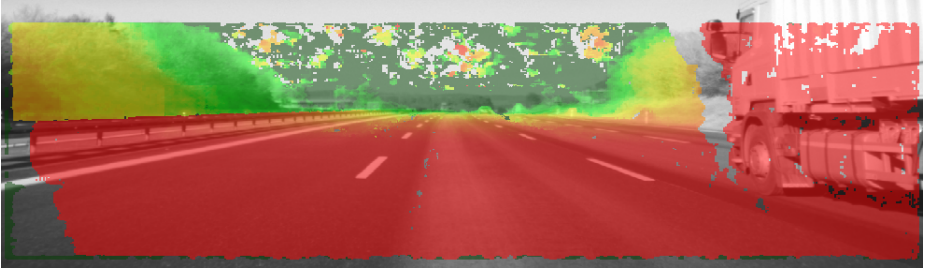


Figure 1.2: Depth information obtained from SGM. Red indicates close distances and green indicates far distances. If no color is assigned to a pixel SGM was unable to assign a distance.

face the camera with a given 3D position and height. The stixel world representation reduces the amount of data significantly in comparison to the results given by SGM. In a road scene scenario, the approximation of having all objects represented as planar objects perpendicular to the ground is also reasonable, i.e. information of interest is retained.

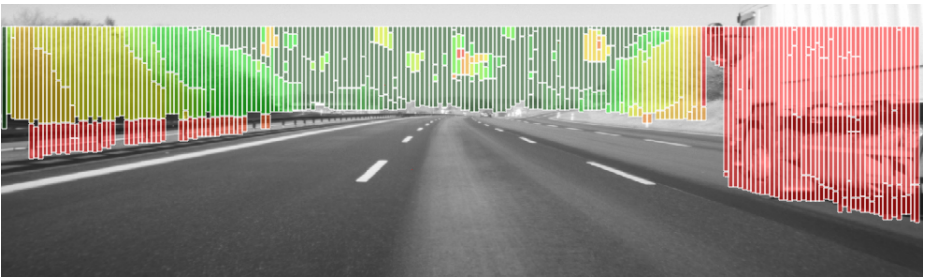


Figure 1.3: The stixel world. Red stixels represent close distances and green stixels represent far distances.

The stixel world provides additional possibilities for modeling the world in order to achieve better results. In [20] gravity and ordering constraints are introduced. Gravity constraints ensure that flying objects are unlikely and that ground adjacent objects should stand on the ground. Ordering constraint ensures that the upper of two staggered objects is expected to be further away from the camera than the lower one.

Limitations of the Stixel World

The stixel world does however have problems with interpreting the environment accurately at large distances. For a practical use of autonomous driving, a vehicle should be able to drive at velocities corresponding to the speed limits of the highways today. The higher the velocity the longer time is required to stop a vehicle, giving larger stopping distance. This implies that the autonomous vehicle must

have an accurate environment perception at large distances. The stixel world fails to interpret the environment at large distances because of the presence of noise and oversmoothing occurring in the stereo disparity computation of SGM. The larger the distance, the lower signal to noise ratio is obtained in the disparity image (see Figure 1.4 (a)). The stixel world also have a fixed road plane assumption which reduces the performance. In other words, one fixed planar road for the entire driving space is assumed.

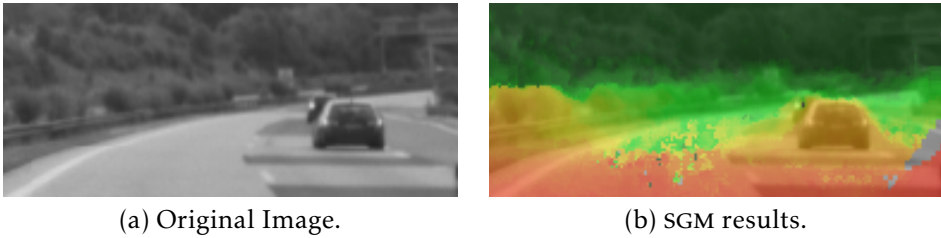


Figure 1.4: In the SGM results at larger distances, areas that should have a difference in depth are assigned very similar values, for instance the regions on both sides of the car in (a) have rather homogeneous values in (b). This is due to oversmoothing in SGM. The presence of noise in the road surface is clearly visible in the center part of (b).

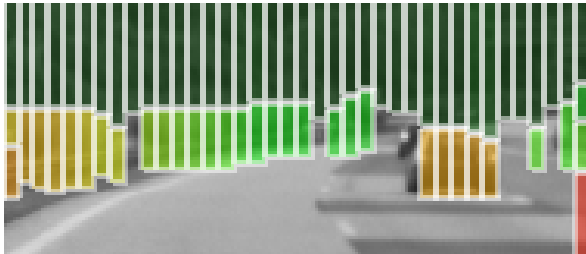


Figure 1.5: Stixel World at larger distances. As can be seen valuable information is missing in the stixel representation, including an insufficient freespace estimate and a undetected occluded vehicle.

As can be seen in Figure 1.4 (b), performing SGM at large distances gives insufficient results. Using SGM results considered as non-satisfying as input data to create stixels is of course resulting in a poor stixel world (see Figure 1.5).

Obtaining spatial perception at large distances, or *long-range road scenes*, is however not an easy problem to solve. For example, long-range road scene perception can be attempted with dense depth estimation and object extraction, i.e the stixel world [3], or appearance based detection of known object classes [7][9]. There are however problems with both these methods. First, as shown above, in dense depth measurement for a scene at a large distance the signal to noise ratio

is very low and problems with oversmoothing occur. The second method uses known object classes, i.e it needs to be trained on the objects before recognizing them. Having known object classes for all possible objects that can appear on the road is however not possible. For instance there can be animals or lost cargo on the road which has no typical appearance. Therefore, there is a need for an algorithm which distinguishes all possible objects in the scene and is able to function in the presence of noise. To achieve this goal, this thesis considers an approach which estimates geometric primitives for all visible elements of a road scene and by applying geometric constraints assigns pixels in the observed image to their respective geometric primitive. This is a labeling task (and thereby detection) of generic objects in a road scene scenario and the aim is to separate the generic objects from freespace, i.e drivable road. In this thesis the main focus is on the labeling task and the integration of geometric constraints.

There are two main difference between the method presented in this thesis and the stixel world. First, in the image domain the methods presented in this thesis enable pixel accuracy and the shape of objects is not limited to rectangles. Second, in the three dimensional case it is possible to obtain slanted surfaces, i.e the geometric primitives are not forced to be facing the camera.

1.1.3 Aim

The goal of the thesis is to perform pixel-wise labeling of gray-scaled images representing long-range road scenes. In more detail, to provide an algorithm for matching pixels to geometric primitives estimated in stereo image pairs. A geometric primitive, or *planar scene element* (PSE), is in this scenario and throughout the report a plane with a given rotation at a given distance from the camera which has a normal vector perpendicular to the normal vector of the ground. The ground is also estimated from the stereo image pairs (see Figure 1.7). Ideally, each PSE represents one of the objects present in the road scene. Sky and ground is to be represented in this way. The labeling of pixels is part of a larger pipeline including the estimation of geometric primitives, i.e 3D plane parameter estimation, and the segmentation (and thereby detection) of objects.



Figure 1.6: Original image, highway road scene.

The first step to be performed in the labeling task is an extraction of a *region of interest* (ROI) in the image. The ROI is in this scenario the part of the image to be considered. Restricting to a ROI is important since it reduces the number

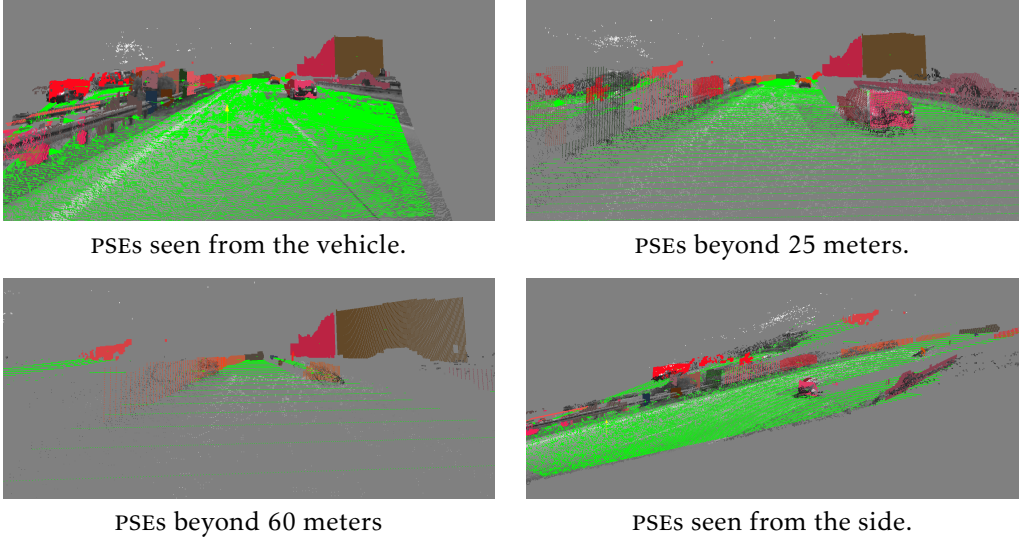


Figure 1.7: Planar scene element representation of the original image in Figure 1.6. The PSE representing ground is green and the PSEs representing sky are excluded in the figures above.

of pixels to label and it creates the possibility to focus on the part of the image representing long-range road scenes.

To compute the labeling, first an objective function (energy function) to be minimized is formulated. It includes likelihoods for each pixel to correspond to an estimated PSE, based on warping and matching of intensities between the stereo image pairs. Moreover, geometric constraints between PSEs, as well as image data such as intensity gradients, are considered. Figure 1.7 illustrates how a scene can be segmented and represented by a set of PSEs.

1.2 Problem

In this thesis a new method for segmenting an image in a man made environment is investigated. To reach the goal of the thesis the following problems are addressed:

- Is it possible to perform image segmentation of a road scene scenario by estimating the scene with geometric primitives and adjusting these?
- Is it possible to obtain results comparable to state-of-the art segmentation algorithms?
- Can a restricted scene model be applied which limits the solution space?
- Can different inference methods be applied?

- Can a real time implementation be achieved?

1.3 Delimitations

Two different kinds of image segmentation methods are investigated in this thesis. Out of these a solution based on graph cuts is investigated more thoroughly. This is because a working limited scene model is prioritized and is easier to incorporate in a graph based solution. Because of this delimitation a graph based method is the only method considered when comparing results obtained in this thesis to other segmentation methods in Section 5.3.

In the presented methods there is no learning incorporated in the solutions. For instance the likelihood of having obstacles present directly in front of the vehicle are rather unlikely and crash barriers have a typical appearance. Theoretically, using learning can improve the solutions but has not been prioritized in this thesis.

The environment investigated in this thesis is highway scenarios. The reason behind this is that the goal is to increase the distance to where objects are first detected and highways are well suited for this purpose.

1.4 Contributions

This thesis investigates how the described piecewise planar scene model and corresponding geometry cues can be used to improve image segmentation and object detection methods in a long-range road scene scenario. The estimation of 3D plane parameters and the computation of pixel-wise likelihood values for plane assignment (data likelihood) was provided by an external algorithm. To reach the goal of and to address the problems of the thesis the following contributions were made. First, a literature study was conducted to obtain knowledge about the current research in the area of the thesis. After the literature study generic labeling algorithms on pairwise Markov Random Fields and Conditional Random Fields was performed. In other words, applying current state-of-the-art labeling algorithms on long-range road scenes for this specific labeling task. This stage included to introduce certain world assumptions and geometric constraints, i.e constraints which enforces certain relations between PSEs forming the resulting labeled image. Creating and investigating the world assumptions and geometric constraints and incorporating them in the labeling algorithms is the main focus of this thesis.

To be able to answer if it is possible to apply different inference methods to this labeling task a dynamic programming approach is investigated. My contributions in this approach is to apply the data available for this labeling task in a dynamic programming method.

Besides performing the image segmentation I contributed in creating the data set

used for evaluation and generating performance measures by comparing results with the generated ground truth data set.

1.5 Report Outline

Chapter 1 explains the need for an algorithm to obtain spatial perception at large distances. It summarizes the work done and aim of the thesis.

A theoretical background is given in Chapter 2 which provides the necessary background information regarding the thesis. It presents and explains representations, methods and algorithms which are included in the thesis.

Chapter 3 presents how an objective function (energy function) can be formulated and thereby formulates the segmentation as an optimization problem. It will present how the problem can be addressed in a more theoretical point of view.

The energy presented in Chapter 3 is on a high level presenting the energy to minimize in order to obtain accurate results. Chapter 4 presents in detail the design choices and restricting world assumptions incorporated in the thesis.

Chapter 5 presents the obtained results from the proposed method. The results are then evaluated in Chapter 5.3. In Chapter 6 the results of the thesis are discussed and a glimpse of future work in the area of the thesis is given. In Chapter 6 the stated questions in Section 1.2 are addressed.

Appendix A gives an alternative dynamic programming based solution to the problem addressed in the thesis.

2

Theoretical Background

The goal of computer vision research is to enable a machine to make predictions about the world, often referred to as visual perception through the process of digital signals [26]. Visual perception is however an inverse problem. We seek to recover unknowns given insufficient information to fully specify the solution. Therefore physics based and probabilistic models are used to disambiguate between potential solutions [21]. Mathematically, visual perception can be formulated as the mapping of the observed data to a latent parameter which correspond to a mathematical answer. Roughly, this can be seen as an optimization problem where the energy function is a quality measure of the solution given the observed data and some parameter assigned to the model. In this specific labeling problem the best labeling is selected from a set hypotheses (any combination of the estimated PSEs) where the task is to assign each pixel a PSE from a finite set of elements. Visual perception is said to include three main tasks: *modeling*, *inference* and *learning* [26]. Modeling includes the task of how to model the real world into a representation which can be interpreted by a computer. Inference is the task of minimizing the energy defined on the model. Learning is the task of how to learn a system to recognize certain patterns in order to improve the solution. In this thesis there is no focus on the learning task.

2.1 Stereo Vision

Stereo Vision is the concept of having two cameras at slightly different positions with a known relative displacement to each other. Scene points captured at the same time by the two cameras are then projected onto the two image planes (see Figure 2.1). Using epipolar geometry, image points corresponding to the same scene point can be found by solving the correspondence problem. The displacement of image points corresponding to the same scene point is referred to as

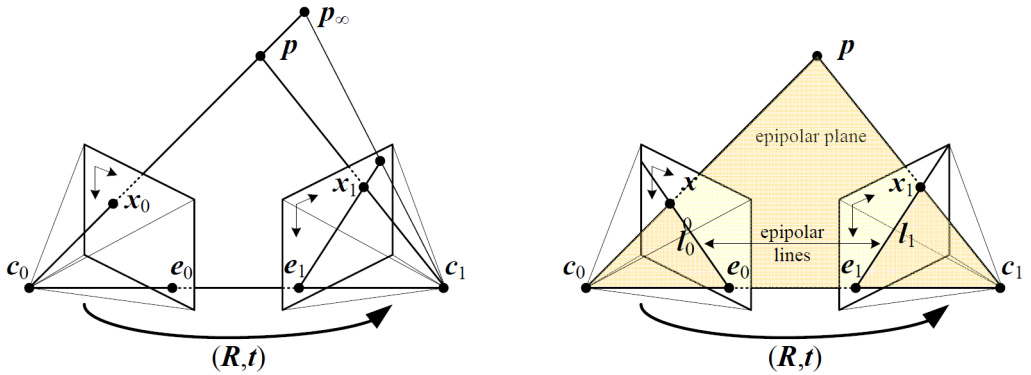


Figure 2.1: Simplified stereo vision system [21].

disparity and the magnitude of disparity values is decreasing with the distance to the scene point. Disparity values are of high interest since they enable triangulation of 3D points. There are two main concepts of disparity estimations from stereo image pairs, local and global methods [24].

2.1.1 Local Methods

Local methods find disparities between points in stereo image pairs by aggregating a matching cost over defined windows and computing disparity of each pixel independently. Local methods are based on pixel similarities (correlation, descriptor matching etc.) and can have efficient implementations suitable for real-time use [15]. The main problem with local methods is that they fail in regions with low texture, i.e if not provided with sufficient data support results will diverge arbitrarily. Accurate matching costs are thereby only available at certain image points.

2.1.2 Global Methods

Global methods optimize over the entire image and are therefore not as dependent on the correlation windows as in the case of local methods (windows can be applied to compute and compare costs in the optimization). A global method is SGM, which generates accurate dense stereo matching [16]. Because of the good trade-off between robustness, accuracy and speed of SGM it is widely used in computer vision applications.

Semi-Global Matching

SGM [13][16] is a global method which performs pixel-wise matching based on a pixel-wise cost calculation and the approximation of a global smoothness con-

straint. The reason why a global smoothness is introduced is because the pixel-wise cost calculation is generally ambiguous and wrong matches can have a lower cost than correct matches. The reason for this can be the presence of noise in the images and image regions with low texture. The additional smoothness term penalizes changes of neighboring disparities. This energy is defined according [16]:

$$E(D) = \sum_p C(p, D_p) + \sum_{q \in N_p} P_1 T[|D_p - D_q| = 1] + \sum_{q \in N_p} P_2 T[|D_p - D_q| > 1] \quad (2.1)$$

where D is the disparity image, q are pixels in the neighborhood N_p of p . $T[\cdot]$ is the probability distribution of corresponding intensities, which is one if its argument is true and zero otherwise. P_1 adds a penalty when the disparity changes by one pixel and P_2 adds a larger penalty for larger disparity changes. Two penalty costs are applied since P_1 allows a better approximation of slanted surfaces and P_2 enables the method to preserve discontinuities in the image. To preserve discontinuities in the image P_2 is adapted to intensity differences in the image. One example of SGM results can be found in Figure 1.2.

2.2 Probabilistic Graphical Models

To model an image for image segmentation, probabilistic graphical models can be used. Probabilistic graphical models are probabilistic models for which a graph denotes the conditional dependence structure between random variables. This form of representation is widely used in the computer vision field. In a probabilistic graphical model, vertices represent random variables and edges represent the conditional dependencies between the vertices it connects. In other words, each vertex represent a pixel to label and the edges represent how the value of a pixel should affect the pixels which are connected to the same vertex. The edges in a graphical model can be directed or undirected, that is the conditional dependence between two variables may only affect one of them. In an undirected graphical model, a subset of vertices where all vertices in the subset is connected by an edge is defined as a *clique*. The most common types of graphical models are Bayesian Networks (BNs) and Markov Random Fields (MRFs) [26].

2.2.1 Factor Graphs

MRFs and BNs have a unified representation called *factor graphs*. The factor graph representation introduces factors to represent potentials assigned to vertices and the conditional dependencies between them. The factor graph was introduced since it provides a more fine-grained representation of the factors that make up the conditional dependencies in a graphical model. The factor graph is

also useful for visualization and most importantly the simplicity of defining inference algorithms on the graph [26]. Figure 2.2 illustrates a simple factor graph representation of a MRF. In computer vision, and especially in the area of labeling pixels, the MRF is widely used.

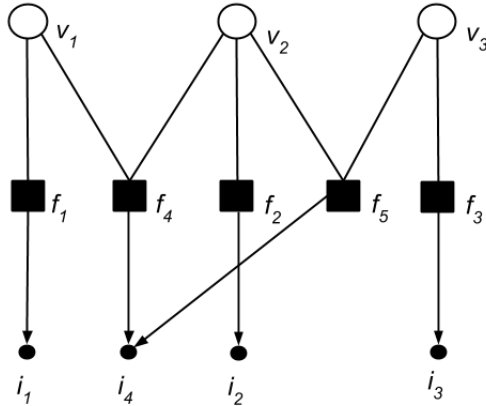


Figure 2.2: A factor graph representation of a MRF with three vertices and two edges. The MRF above contains three unary potentials and two pairwise potentials.

2.3 Markov Random Fields

A MRF is a set of random variables, vertices, forming an undirected and possibly cyclic graph which holds the Markov Property. The Markov Property imposes that a random variable in the MRF is independent of any other variables given all its neighbors [26]. A MRF differs from the BN which is directed and acyclic. In this thesis there is no focus on BNs.

The MRF is widely used in computer vision applications, mainly because of the strengths of the MRF properties. This includes the simplicity of combining different likelihood terms and other useful data within a single graph representation, a simple way of visualizing the model and factorization of the joint probability over the graph which gives inference problems that can be solved efficiently [26]. In favor of simplicity and computational efficiency, the most common type of MRF for computer vision applications is the *pairwise* MRF. In a pairwise MRF the energy is factorized into a sum of potential functions which are defined on cliques with an order of strictly less than three. That is, pairwise MRFs can be represented by a graph containing unary potentials assigned to the random variables and a set of pairwise potentials assigned to pairs of random variables within the graph [26].

The energy of a MRF can be derived from *Bayes' Rule*. The posterior distribution for a set of measurements y , $p(y|x)$, combined with a prior $p(x)$ obtained

from the unknowns x , is given by [21]:

$$p(x|y) = \frac{p(y|x)p(x)}{p(y)} \quad (2.2)$$

where the denominator is a constant ensuring a proper distribution. The MRF is used to model the prior distribution $p(x)$. The prior distribution is also a Gibbs distribution [26]. Taking the negative logarithm on both sides of Equation 2.2 gives the *negative posterior log likelihood*:

$$-\log p(x|y) = -\log p(y|x) - \log p(x) + C \quad (2.3)$$

To find the maximum likelihood of (2.3) the negative log likelihood is minimized. The constant C is neglected since its value has no effect during minimization. The entire energy of the MRF representing an image can therefore be expressed as:

$$E = \sum_{p \in \Omega} D_p(l_p) + \lambda \sum_{(p,q) \in \mathcal{N}} V(l_p, l_q) \quad (2.4)$$

where $D_p(l_p)$ is the data cost (given by the unary potential) for assigning label l_p to pixel p and $V(l_p, l_q)$ is the label inconsistency cost, or smoothness term (given by the pairwise potential). \mathcal{N} is a neighbourhood defined by the pairwise potentials assigned in the MRF. The pairwise potentials represents the cost for assigning certain labels to two vertices forming a pair in the MRF.

2.3.1 Conditional Random Fields

Similar to MRF modeling, one can also use *Conditional Random Fields* (CRFs). A CRF uses a conditional distribution over the latent variables which gives a more flexible way of incorporating observed variables. The Bayesian derivation of (2.2) does not hold for the CRF, since a CRF describes not only the prior but the complete distribution [21]. By using a CRF, pair potentials can be dependent on the input data itself. CRFs are used in this thesis since intensity differences in the image are used when providing pair potentials.

2.4 Modeling MRFs

The most common graph structure for computer vision applications based on MRFs is the pairwise grid structure (consider a 4- or 8-connectivity pixel neighbourhood, see Figure 2.3). In a grid structure each vertex represents a pixel in

the image and edges represent which pixels that should affect another while inferring the model. In order to improve quality of results (not considering time usage) one might resort to using *higher-order* MRFs, which contains potential functions defined on cliques of order larger than two. The main benefits of this is the possibility to model more complex and natural statistics and allow a richer interaction between vertices. The downside of allowing this is the increment of complexity in inference methods. Higher-order MRFs are not investigated in this thesis because of this, inference methods are considered complex enough using pairwise MRFs.

In pairwise MRFs potentials can be assigned on single random variables (unary potentials) and to pairs of random variables (pairwise potentials). A questions often addressed when designing a MRF/CRF is how to assign these potentials. There is no clear answer regarding this since different applications often require different potentials assigned to different pairs of random variables. The design of the MRF/CRF is however of great importance since it will affect the outcome and complexity of both the construction of the model and the inference algorithm minimizing the defined energy.

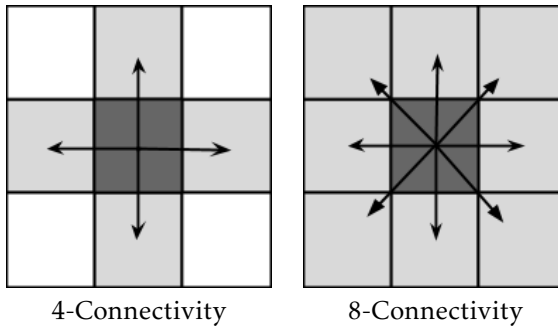


Figure 2.3: Common pixel neighborhoods.

2.4.1 Potts Model

A simple but yet widely used and powerful method for assigning pair potentials is the *Potts model*. The Potts model is defined as:

$$V(l_p, l_q) = \omega \cdot (1 - \delta(l_p - l_q)) \quad (2.5)$$

where the pairwise potential $V(l_p, l_q)$ between the pixels p and q and labels l_p and l_q corresponds to the value of a weight, ω , if the labels l_q and l_p differ. If the labels are the same the pairwise potential is set to zero. In other words the Potts model penalizes changes of labels between vertices connected by an edge.

2.4.2 User Defined Models

One of the strengths of MRFs is the simplicity of combining different likelihood terms within the same graph representation. The simplicity comes from the fact that the MRF model can be specified by the simple summation of the included potentials. All valid available data can therefore be applied. For instance, PSEs contain valuable depth information about the probabilistic whereabouts of each pixel. This information can therefore be included in the model and give the possibility to obtain better results. However, since (for all but the simplest models) it is very hard to directly derive these potentials themselves from probabilistic measures, the problem with the correct *scaling factors* arises. In other words, applying data retrieved from different sources is creating the issue of finding a good trade-off between them.

2.5 Inference Algorithms on Pairwise MRFs

Statistical inference is the idea of drawing conclusions from data that is affected by random variation. Computer vision applications based on MRFs or CRFs have the essential problem of how to infer the optimal configuration for all vertices. This problem is found to be NP-hard in general for the multi-class labeling problem [26][21]. Inference algorithms performed on MRFs want to find the minimum of equation 2.4:

$$E_{min} = \min \left\{ \sum_{p \in \Omega} D_p(l_p) + \lambda \sum_{(p,q) \in \mathcal{N}} V(l_p, l_q) \right\} \quad (2.6)$$

Doing this for the entire set of Ω (all pixels in the image or defined ROI) will infer the entire graph and provide a solution. There are three main classes of inference methods used today for pairwise MRFs and CRFs. This includes *graph cuts*, *belief propagation algorithms* and *dual methods* [26]. These three are used since they are powerful in practice. In this paper there is no focus on dual methods or belief propagation, mainly because of the popularity and strengths of graph cuts. Inference methods based on graph cuts depend on initialization. A different initialization of the inference method can result in a different solution for the multi-class labeling problem.

2.5.1 Graph cuts

Inference using graph cuts is based on the idea to form a directed graph, called *s-t graph*, which contains two special *terminal* vertices. The terminal vertices are called *sink* and *source* and the cut that is to be made must separate these vertices. The cut is also determining the label for vertices in the graph based on if the edge connecting the vertex is included in the cut. Figure 2.4 illustrates this. For larger graphs there exists many possibilities of cutting the graph which

separates the terminals, leading up to the problem of finding the optimal cut. This is referred to as the *minimal cut problem*. The minimal cut problem finds the cheapest cut cost among all cuts separating the terminal vertices. A cut cost is defined as the sum of the edges within the cut, i.e. the sum of the weights of the edges removed to separate the sink and the source [5]. A MRF or CRF which has such an s-t graph is called graph-representable and the minimization of the energy of the MRF or CRF is equivalent to solving the minimal cut problem [26]. An optimal cut can be found between two labels. For the multi-labeling problem this is not the case. Graph cuts can thereby only give solutions which approximates the optimal labeling of the graph. The two most common algorithms for performing multi-label graph cuts is α -expansion and $\alpha\beta$ -swap. Both methods are iterative move making algorithms. They optimize the MRF energy by defining a set of possible moves based on initialization values and sets the best move as initial configuration for the next iteration. Move-making algorithms run until convergence or until a maximum number of iterations have been reached [26].

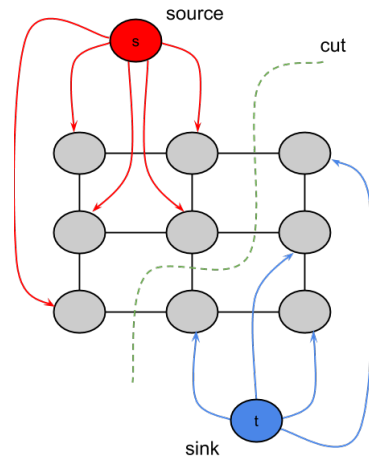


Figure 2.4: Illustration of a graph cut.

$\alpha\beta$ -swap

$\alpha\beta$ -swap starts from initial labeling and for each vertex it finds if the label should be assigned to label α , β or remain to the initial label. This decision is based on where the cut is made within the s-t graph. If the cut is containing the edge connecting the terminal node representing the α label and not the terminal node representing the β label, the vertex in question is assigned the label α . Vice versa applies for the β label. If the cut is not following any of the patterns above, the vertex in question will remain the same label [5]. In other words it iterates over all pairs of distinct labels α and β and, in each iteration, a binary problem is constructed based on the question which vertices that are currently labeled

α should be labeled β such that the improvement over the current labeling is optimal [2].

α -expansion

α -expansion starts from initial labeling and creates the s-t graph with one terminal node representing the label α and one terminal node representing the initial label given to the vertex in question. As in the case of $\alpha\beta$ -swap it will assign the label α if the cut contains the edge between the terminal node representing the α -label and not the vertex in question. Vice versa applies for the terminal node representing the initial label [5]. In other words in each iteration, a problem with binary variables is constructed based on the question of which subset of vertices whose current label is not α should be labeled α to gain an optimal improvement with respect to the current labeling [2]. Note that $\alpha - \beta$ -swap is more general, but α -expansion is faster and gives better results in practice. The complexity of α -expansion is linear and $\alpha - \beta$ -swap is quadratic in the number of labels.

2.5.2 Submodularity

In computer vision community there is an accepted view that graph cuts can only be used to minimize submodular energy functions [17]. To be considered submodular, pairwise potentials defined on binary models must satisfy the following condition:

$$V_{pq}(\alpha, \alpha) + V_{pq}(\beta, \beta) \leq V_{pq}(\alpha, \beta) + V_{pq}(\beta, \alpha) \quad (2.7)$$

for all possible values of the labels α and β . E_{pq} is the pairwise potential applied to the edge between vertex p and q . For a multi-class labeling problem, and specificity for $\alpha - expansion$, the pairwise potentials must satisfy:

$$V_{pq}(\alpha, \alpha) + V_{pq}(\beta, \gamma) \leq V_{pq}(\beta, \alpha) + V_{pq}(\alpha, \gamma) \quad (2.8)$$

for all possible potential values of the labels α , β and γ to be considered submodular [17]. For instance, the Potts model is fulfilling this condition. More complex energy functions can however cause the condition above to fail and the designer must therefore be careful when modeling the graph.

In [23] pairwise potentials which are non-submodular are truncated in order to obtain submodularity. That is, one of the pair potentials in Equation 2.8 is decreased/increased until the condition is valid. [23] also proves that it is possible to include hard constraints, i.e E_{pq} can take values in $\{0, +\infty\}$ as long as $E_{pq}(\alpha, \alpha) = 0$. Hard constraints can however provide pair potentials which are non-submodular. Pair potentials representing infinity can be non-submodular but as long as they never appear in the solution it will not cause a problem.

2.6 Limitations of MRFs

MRFs and CRFs have their benefits and strengths. Using MRF/CRFs for certain applications does however contain disadvantages. This is due to the fact that in some applications there is no guarantee to reach the global optimum while inferring. Binary-labeling problems are, with the right optimization algorithm, guaranteed to produce a global optimum. The multi-class labeling problem is however not, as the problem is found to be NP-hard [21]. Solving the multi-class labeling problem must therefore require algorithms which approximate the global optimum.

Assigning a model as simple as the Potts model in (2.5) is seldom enough for solving a more complex problem as the multi-class labeling problem accurately. This is partly because there might exist pixels which should affect the outcome label of a pixel which do not exist in the defined neighbourhood. This is the main difficulty in modeling; many vision problems are inverse, ill-posed and require a large amount of variables to express the expected variations of the answer to the visual perception problem [26]. This implies that in many applications potentials must be extended to not only rely on the neighbouring pixels defined by a 4-connectivity neighbourhood. In some applications the 8-connectivity or even higher order neighbourhoods perform better at tasks such as image segmentation because they can better model discontinuities at different orientations [21]. Higher order potentials could model more powerful dependencies (e.g. dependence of one pixel label on a whole region), but inference is in general harder. A problem with extending the number of pair potentials is that there is no distinguished level where the number of pair potentials are considered sufficient. Ideally when assigning a label to a vertex in a MRF or CRF the best case would be to consider all other vertices in the graph. Doing this for all vertices is not possible in the case of representing large images as a graph since the number of vertices would be high, giving computationally expensive inference algorithms. This means that there exists a trade-off between having a fast inference algorithm and how exact the model approximates the scene. If MRFs or CRFs are used when modeling the multi-class labeling problem it must be possible to model the scene accurate enough to obtain results considered sufficient and fast enough. Because of this there will always exist a limitation in the energy function; potentials defined will affect the result and run time of the inference algorithm. This restriction implies that using pixel-wise MRFs or CRFs to find the best model for long-range road scenes may be challenging considering a real time implementation. Once the order of the potentials and the connectivity is chosen, the problem of assigning suitable values still exists. Problems such as maintaining submodularity needs to be dealt with to be able to find a solution.

3

Energy Formulation

The energy defined on MRFs presented in (2.4) explains how the energy function in a MRF is formulated and what it may contain. This chapter intends to give a deeper understanding on how the energy to minimize in this optimization problem is formulated. The work is described as a joint labeling and parameter estimation task.

3.1 Joint Labeling and Parameter Estimation

Considering the following notation:

$\vec{\mathcal{I}}$... Set of input data (Intensity measurements of the stereo image pair).

$\vec{\mathcal{L}}$... Discrete random variables representing the labeling of each image pixel in the reference image.

$\vec{\ell}$... Realization of $\vec{\mathcal{L}}$.

$\vec{\omega}$... Continuous random variables representing the set of parameters of all K scene elements (plane parameters: three parameters holding the plane normal and one parameter holding the distance from the origin).

\vec{w} ... Realization of $\vec{\omega}$.

The posterior probability distribution of the labeling and the parameters can be described given the observed measurements using Bayes' rule:

$$p(\vec{\ell}, \vec{w} | \vec{\mathcal{I}}) = \frac{p(\vec{\mathcal{I}} | \vec{\ell}, \vec{w}) p(\vec{\ell}, \vec{w})}{p(\vec{\mathcal{I}})} = \frac{p(\vec{\mathcal{I}} | \vec{\ell}, \vec{w}) p(\vec{\ell} | \vec{w}) p(\vec{w})}{p(\vec{\mathcal{I}})} \quad (3.1)$$

What is searched for is the realizations of $\vec{\mathcal{L}}$ and $\vec{\omega}$ that maximize $p(\vec{\ell}, \vec{w} | \vec{\mathcal{I}})$, i.e.

the maximum-a-posteriori (MAP) estimates $\widehat{\vec{\ell}}$ and $\widehat{\vec{w}}$.

During optimization the denominator $p(\vec{\mathcal{I}})$ remains constant and has no influence on the result. This means that what is searched for is:

$$(\widehat{\vec{\ell}}, \widehat{\vec{w}}) = \arg \max (p(\vec{\mathcal{I}}|\vec{\ell}, \vec{w})p(\vec{\ell}|\vec{w})p(\vec{w})) \quad (3.2)$$

The distribution in (3.2) consists of three terms: the *data likelihood*, $p(\vec{\mathcal{I}}|\vec{\ell}, \vec{w})$, the *label prior*, $p(\vec{\ell}|\vec{w})$, and the *parameter prior*, $p(\vec{w})$.

3.1.1 Formulation of the Different Terms

Data Likelihood

The data likelihood $p(\vec{\mathcal{I}}|\vec{\ell}, \vec{w})$ is modeled using the observed intensities of the stereo image pair. Corresponding points in the two images is assumed to contain equal intensities (brightness constancy assumption). Furthermore, a simple image noise model of Gaussian noise (independent and identically distributed for each pixel) is assumed. Given these assumptions, a model for $p(\vec{\mathcal{I}}|\vec{\ell}, \vec{w})$ using the pixel intensity differences can be stated:

$$\log (p(\vec{\mathcal{I}}|\vec{\ell}, \vec{w})) \propto \sum_{k=0}^{K-1} \sum_{p \in \Omega_k} (I_l(p) - I_r(f(p, \vec{w}_k)))^2 \quad (3.3)$$

where I_l and I_r are the left and right stereo images, Ω_k is the pixel support of the scene element k (according to the labeling $\vec{\ell}$) and the function f represents the warping of the coordinates of pixel p from the left into the right image, according to the parameters \vec{w}_k .

Note that this is a quite simple model which is vulnerable to violations of the brightness constancy assumption. However, since the aim of the thesis is to investigate improvements from introducing constraints on the labeling this is not crucial.

Parameter Prior

For the parameter prior uninformative (i.e. uniformly distributed) priors on the parameters is assumed, which do not influence the result of the estimation.

Label Prior

The label prior is the focus of this work and there is a difference in the formulation for a MRF and a CRF.

In the MRF case, Bayes' rule can directly be applied and split up the different terms as in (3.2). The MRF then only models the prior term $p(\vec{\ell}|\vec{w})$ (using the binary potentials).

In the CRF case, the labeling is dependent on the input data $\vec{\mathcal{I}}$ (e.g. image intensity etc. is considered). Therefore, the terms *cannot* be split up as in (3.2) anymore and have to model the posterior $p(\vec{\ell}, \vec{w} | \vec{\mathcal{I}})$ directly using a CRF (see [21] on MRFs/CRFs). The data likelihood term $p(\vec{\mathcal{I}} | \vec{\ell}, \vec{w})$ then simply appears as a unary potential in the CRF.

The MRF/CRF describes the respective joint distribution of the graph as a Gibbs distribution (Hammersley-Clifford Theorem [14], also [26]). For the CRF, it is of the form:

$$\frac{1}{Z(\vec{w}, \vec{\mathcal{I}})} \prod_{c \in \mathcal{C}} \psi_c(\vec{\ell}_c, \vec{w}, \vec{\mathcal{I}}) = \frac{1}{Z(\vec{w}, \vec{\mathcal{I}})} \exp(-E(\vec{\ell}, \vec{w}, \vec{\mathcal{I}})), \quad (3.4)$$

where Z is the partition function (a normalizing factor) and $\psi_c(\vec{\ell}_c)$ is the potential function of the clique c (holding the subset of variables $\vec{\ell}_c$ of $\vec{\ell}$).

The energy E can be written as a sum of clique potentials V_c :

$$E(\vec{\ell}, \vec{w}, \vec{\mathcal{I}}) = \sum_{c \in \mathcal{C}} V_c(\vec{\ell}_c, \vec{w}, \vec{\mathcal{I}}) \quad (3.5)$$

where $V_c(\vec{\ell}_c, \vec{w}, \vec{\mathcal{I}}) = -\log(\psi_c(\vec{\ell}_c, \vec{w}, \vec{\mathcal{I}}))$. For pairwise CRF models, the associated energy and thus the joint posterior distribution can be specified by unary and binary potentials:

$$\begin{aligned} E(\vec{\ell}, \vec{w}, \vec{\mathcal{I}}) &= \sum_{p \in \mathcal{C}_1} V_1(\ell_p, \vec{w}, \vec{\mathcal{I}}) + \sum_{p, q \in \mathcal{C}_2} V_2(\ell_p, \ell_q, \vec{w}, \vec{\mathcal{I}}) \\ &= \sum_{p \in \mathcal{S}} V_1(\ell_p, \vec{w}, \vec{\mathcal{I}}) + \sum_{p \in \mathcal{S}} \sum_{q \in \mathcal{N}_p} V_2(\ell_p, \ell_q, \vec{w}, \vec{\mathcal{I}}). \end{aligned} \quad (3.6)$$

where \mathcal{C}_1 and \mathcal{C}_2 are cliques of order one and two and \mathcal{S} is the set of all vertices in the graph.

3.1.2 Inference in Practice

In practice a truly joint estimation of $\vec{\ell}$ and \vec{w} *cannot* be performed. Therefore an iterative approach, computing alternating updates to $\vec{\ell}$ and \vec{w} is chosen.

While updating the labeling $\vec{\ell}$, the current estimate of the parameters \vec{w} is held fixed, which means that also the partition function $Z(\vec{w}, \vec{\mathcal{I}})$ remains constant and can be ignored in this optimization step.

When updating the parameter estimate \vec{w} , the labeling $\vec{\ell}$ is held fixed. To guarantee a continuous increase of the target function derived from (3.1), the computation of \vec{w} also needs to take $p(\vec{\ell} | \vec{w})$ into account. Note that now also the partition function Z varies with the parameters and would have to be evaluated. In practice this is intractable, but an approximation e.g. by the pseudo-likelihood could be used [19]. However, for simplicity the term $p(\vec{\ell} | \vec{w})$ is not included in

the implementation of the parameter estimation in this thesis, meaning that the *coupling* between labeling and parameters is only enforced in the labeling step. The plane parameter update is therefore done by directly minimizing (3.3) via Gauss-Newton iterations. On the one hand this might cause an increase of the energy of the labeling during the parameter estimation step, but on the other hand the overall solution is less likely to get stuck in local optima. Note that this simplification can be a reason for what is modeled as a hard constraints not always remain enforced, i.e results can contain PSE relations found as unlikely.

Label Costs

In order to restrict the solution *label costs* can additionally be introduced [8]. Label costs are not dependent on any neighborhood or PSE combination defined on a clique. This energy can be included in order to restrict the complexity of solution in certain ways, for instance to reduce the number of scene elements. Label costs can be added when a new PSE is included in the solution, thereby penalizing the number of PSEs used. In other words, label costs are introduced to explain the data with fewer, cheaper labels [8]. For simplicity, label costs are approximated by a constant for all new labels.

4

Implementation

The methods presented in this thesis can be modelled in various ways and results can be obtained with many different parameter settings. This chapter will further present the data and information included and how they are modelled as potentials included in the energy in (3.6).

4.1 Modeling Label Priors

Various costs can be added to the energy function in both a MRF/CRF and a dynamic programming approach of this labeling task. In a long-range road scene scenario, where the raw data is not as strong as for short-range, it is important to exploit the data and information which in fact is available. This means that there is a need to include all valuable information and find a good trade off between them. Additional potentials added to obtain a more accurate description of the world are included as label priors. The priors are extracted from image intensity values and relations between PSEs. This section will further present these priors and how they are modeled as potentials included in the energy function.

4.1.1 Applied Cost Function

The energy function is defined according to (3.6). In order to adjust the impact of the various *pair potentials* and to improve the results by introducing the Potts model the following cost function is applied:

$$V_2(\ell_p, \ell_q, \vec{w}, \vec{I}) = \begin{cases} \omega + \lambda \cdot -\log(L_{\text{PSE}}(p, q, l_p, l_q)) + \gamma \cdot -\log(L_{\text{int}}(p, q)), & \text{if } l_p \neq l_q \\ 0, & \text{otherwise} \end{cases} \quad (4.1)$$

where ω is the Potts strength, λ is a factor balancing the influences of label priors given by relations between the considered PSEs (L_{PSE}), γ is a factor balancing the influences of image intensity differences (L_{int}). p and q are the pixels considered and l_q and l_p are the PSEs considered. In addition the the applied cost function, unary potentials and label costs are included (as described in Section 3.1.2) in the energy function.

4.1.2 Relations between PSEs

Based on the estimated plane parameters, depth relations between PSEs can be included in the model applied in the MRF/CRF approach and hence give a better approximation of the real world. The cost of this is however the need for additional data and a computation of the relations. To achieve a better approximation of the world, additional restrictions are applied based on the difference in depth between neighboring PSEs. There exist three main labels: *sky*, *ground* and *obstacle*. Sky and ground is represented by one label but obstacles can be represented by multiple different labels. The restrictions based on relations between PSEs are similar to ordering and gravity constraints for stixels [20]. In the stixel world, constraints are only applied vertically, in this labeling task there is however possible to apply constraints horizontally. The vertical constraints are however considered as stronger constraints which shows a greater influence on the results. Using depth restrictions between objects is for instance useful to enforce that no object is located further away than the modeled sky/background. Relations between PSEs are transformed into a potential in the energy function by applying different potentials for different combinations of PSE and pixel relations. In this thesis two different types of methods have been investigated when considering relations between PSEs: *approximate hard constraints* and *likelihood functions*.

Approximate Hard Constraints

Potentials modeling approximate hard constraints are intended to prevent labeling configurations considered to be impossible in real-world road scenes, based on the geometry of the PSEs. To this end, potentials representing label-changes either take on a constant (i.e. the Potts strength) or a value representing zero probability (infinite cost). This method is used for obtaining results presented in Chapter 5. One important thing to notice is that even though an assignment of a probability representing zero, i.e a cut-cost representing infinity, does not mean that it will never occur in the solution. Assigning a probability of zero means that a very high costs is assigned while performing the graph-cut algorithm. Having a cut-cost of infinity is however not possible in the current implementation to prevent a data type overflow.

Likelihood Distribution

Alternatively, the computation of the potentials can be motivated heuristically by defining and sampling separate likelihood functions. Truncation is applied

to fulfil the submodularity constraints in (2.8). The potentials are motivated by the likelihood of the depth value of one considered pixel, given a certain labeling, and the depth of the other considered pixel.

The likelihoods applied to the energy function, L_{PSE} , given by approximate hard constraints can be found in Table 4.1 and Table 4.2. Likelihoods given by a defined distribution can be found in Table 4.3.

l_q	l_p	Condition	$-\log(L_{\text{PSE}})$
o	o	$ Z_p - Z_q < \Delta_o$	∞
o	o	$ Z_p - Z_q \geq \Delta_o$	ω
o	s	$Z_p + \Delta_z \geq Z_{max}$	∞
o	s	$Z_p + \Delta_z < Z_{max}$	ω
o	g	$Z_p - Z_q \leq \frac{Z_p^2 \epsilon_d}{bf\epsilon_d + \epsilon_\Delta}$	ω
o	g	$Z_p - Z_q > \frac{Z_p^2 \epsilon_d}{bf\epsilon_d + \epsilon_\Delta}$	∞
g	o	$Z_p - Z_q < 0$	ω
g	o	$Z_p - Z_q \geq 0$	∞

Table 4.1: Vertical potentials applied between PSEs based on difference in depth between them. The potential functions are returning the Potts strength or the likelihood of zero depending on if a threshold is reached. l_p represents the upper PSE and l_q the below. Z is the distance to a certain PSE or maximum/minimum distance, Δ_o is the minimum distance allowed between two objects, Δ_z and ϵ_Δ are small offset value, ϵ_d is disparity errors given in pixels (see Figure 4.2), b is the base of the camera and f is the focal length of the camera. Note that a cut cost of infinity is only approximated.

l_q	l_p	Condition	$-\log(L_{\text{PSE}})$
g	o	$Z_p - Z_q > 0$	∞
g	o	$Z_p - Z_q \leq 0$	ω
s	o	$Z_p - Z_q > 0$	∞
s	o	$Z_p - Z_q \leq 0$	ω
o	g	$Z_p - Z_q < 0$	∞
o	g	$Z_p - Z_q \geq 0$	ω
o	s	$Z_p - Z_q < 0$	∞
o	s	$Z_p - Z_q \geq 0$	ω

Table 4.2: Horizontal potentials applied between planar scene elements based on the difference in depth between them. l_p represents the left PSE and l_q the right.

l_q	l_p	Condition	L_{PSE}
o	o	$ Z_p - Z_q < \Delta_o$	0
o	o	$Z_p - Z_q \geq \Delta_o$	$\frac{1 - p_{ord}}{Z_{max} - Z_q - \Delta_o}$
o	o	$Z_p - Z_q + \Delta_o < 0$	$\frac{p_{ord}}{Z_q - \Delta_o - Z_{min}}$
o	s	$Z_p + \Delta_z \geq Z_{max}$	0
o	s	$Z_p + \Delta_z < Z_{max}$	$\frac{1/(Z_{max} - Z_{min})}{1 - \Delta_z/(Z_{max} - Z_{min})}$
o	g	$\forall Z_q, Z_p$	$\frac{bf}{\sqrt{2\pi}\sigma Z_q^2} \exp\left(-\frac{(\frac{bf}{Z_q} - \frac{bf}{Z_p})^2}{2\sigma^2}\right)$
g	o	$Z_p - Z_q < 0$	0
g	o	$Z_p - Z_q \geq 0$	$\frac{1}{Z_{max} - Z_q - \Delta_o}$

Table 4.3: Vertical likelihoods applied between PSEs based on difference in depth between them. The likelihood functions, L_{PSE} , are given by the depth values. l_p represents the upper planar scene element and l_q the below. Z is the distance to a certain PSE or maximum/minimum distance, Δ_o is the minimum distance allowed between two objects, Δ_z is a small offset value, p_{ord} and σ are adjustable parameters, b is the base of the camera and f is the focal length of the camera. Note that the negative logarithm is to be applied before the values can be included as cost in an energy function. Also note that the values in this table are not used when presenting and evaluating results.

Distances assigned to PSE are obtained from the estimated PSE parameters. In long-range road scenes the distance from the camera is relatively large, giving noisy distance measurements. The large distance considered in long-range road scenes also gives that the differences in depth between two adjacent pixels can be large on slanted surfaces. To reduce the impact of these factors and because it is desired to compare distances where the actual cut is to be made (the actual cut is made at the edge of a pixel) an interpolation is performed before comparing distances between the considered PSEs. The depth difference between two adjacent pixels representing the ground plane is approximately 20 meters at a distance of 175 meters, clearly motivating the need for an interpolation. The interpolation performed is between the depth value of the pixels and the PSEs considered. For simplicity linear interpolation is performed. Linear interpolation is considered to be a sufficient interpolation method for reducing the error caused by this. The following sections explain which combination of PSEs which are considered and why they provide valid restrictions in this labeling task. When referring to *top*, *bottom*, *left* and *right* in the sections below it refers to the relative location of pixels in the image plane between which the edge (to be cut or not) lies.

Ground on top of Sky

The scenario of having ground on top of sky is modeled to be impossible. This is almost always the case for the real world. If there would be a scenario where ground is above sky the ground segment would be assigned to the sky or a object.

Object to the left or to the right of Ground

An object to the left or to the right of ground should appear to be closer than the ground. If this is not the case, the object would appear to be inside the road which, as in the real world, is not considered a possible scenario.

Object to the left or to the right of Sky

An object to the left or to the right of sky should appear to be closer than the sky. If this is not the case, the object would be further away than what is modeled as infinity. In other words objects are not allowed to be further away than sky, which also applies to the real world.

Ground on top of Object

Having ground above object should only apply in the case where the autonomous vehicle is facing uphill or downhill, observing ground above the objects in front of it. This is a scenario which can occur but is not seen as scenario with high probability.

Object on top of Ground

Having object on top of ground is a scenario which applies to all scenes containing an object. The assumption that no objects should be found flying is made,

i.e. objects directly above the ground should always be connected to the ground. A cut between object and ground is therefore more likely to happen when the depth distance between object and ground is close to zero. This constraint enforces that borders between objects and ground are more accurate. In the case of binary potentials given by a distribution the following likelihood function, L_{PSE} , is applied:

$$L_{\text{PSE}} = \frac{bf}{\sqrt{2\pi\sigma Z_o^2}} \exp\left(-\frac{\left(\frac{bf}{Z_o} - \frac{bf}{Z_g}\right)^2}{2\sigma^2}\right) \quad (4.2)$$

where Z_g and Z_o is the distance to ground and object, b and f is the base and focal length of the camera and σ is an adjustable parameter. The distribution for this constraint gives that the closer the object is to the vehicle the higher likelihood can be obtained (see Figure 4.1). This is because the uncertainty of the data increases with the distance from the vehicle.

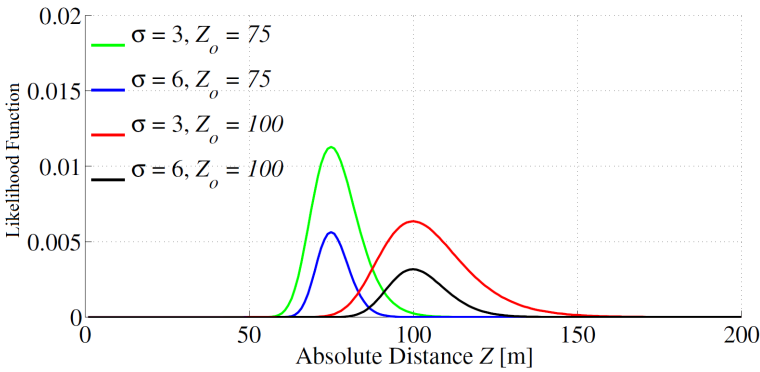


Figure 4.1: Likelihood of a cut between object to ground with specified camera parameters at a object distance of 75 and 100 meters. The distribution is centered around the distance to the considered object and the magnitude is higher with a shorter object distance. The width of the distribution is increasing with the distance in order to allow larger deviations between object and ground where values are affected by noise.

In the case of binary potentials representing approximate hard constraints the decision is made if the distance between object and ground at the position in question is within a certain interval. This interval is defined by the metric distance error (see Figure 4.2), ϵ_z , and a small offset value, ϵ_Δ :

$$\epsilon_z = \frac{Z_p^2 \epsilon_d}{bf \epsilon_d} + \epsilon_\Delta \quad (4.3)$$

where ϵ_d is the assumed disparity error given in pixels and b and f is the base and focal length of the camera. The metric distance error is used since the errors increases with the distance. Therefore it is reasonable to allow a larger deviation at larger distances, which the metric distance error gives. The metric distance error is derived from:

$$\epsilon_z = \frac{bf}{d} + \frac{bf}{d - \epsilon_d} = \frac{Z_p^2 \epsilon_d}{bf \epsilon_d} \quad (4.4)$$

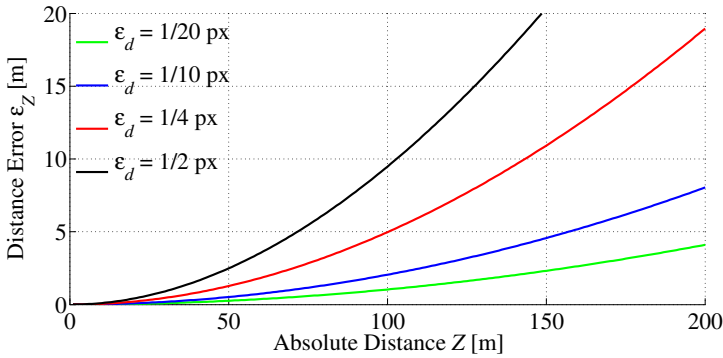


Figure 4.2: Metric distance errors ϵ_z . The distance error increases non-linearly for given stereo disparity errors.

Object on top of Sky

Having object on top of sky is a scenario which applies when there is an object above the road, for instance a bridge or certain traffic signs. These situations can occur but are not modeled as very likely in comparison to others.

Object on top of Object

Having an object on top of another object is a scenario which is common for long-range road scenes, for instance when a large truck is in front of small car. The object which is above should be further away then the object below. If the below object would be further away it should receive a lower likelihood and thereby a higher cost. In this scenario it is also possible to threshold the distance between the considered objects. In the example with a large truck and the small car, the corresponding PSEs are expected to have a distinguished difference in depth. If this is not the case a low likelihood is to be assigned.

Sky on top of Object

Having sky on top of an object is a scenario which applies for almost every scene containing an object. The likelihood for having sky closer than the object is however assumed to be impossible since this would give an object at a distance modeled as infinity.

Sky on top of Ground

Having sky on top of ground is a scenario which applies when there is no object at the considered column in the image. The sky should naturally be located at a further distance than the ground and if this is not the case a high cost is to be applied.

4.1.3 Relations between PSEs and estimated horizon

If the horizon is available advantages appear when modeling the world. For instance, with a given horizon there is a limit for where it is possible to find ground and sky (ground is not possible to be above the horizon and sky is not possible to be below the horizon). An estimate for the horizon can be supplied by an external algorithm or using the previously estimated PSEs representing ground and sky. An example of a horizon estimate extracted from PSEs can be found in Figure 4.3.

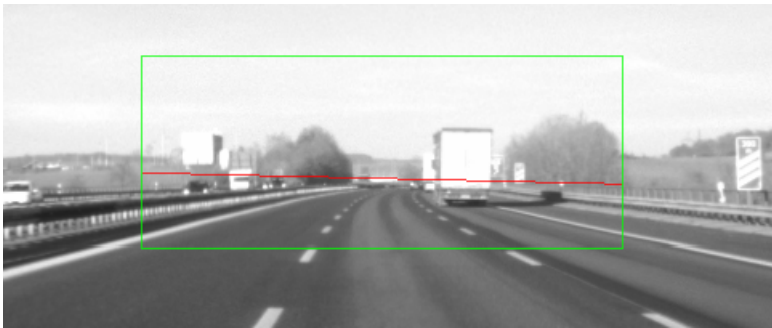


Figure 4.3: *Estimated horizon in specified ROI. The horizon is the red line and the ROI is within the green bounding box.*

If a valid horizon estimate is available, different likelihoods can be applied for certain PSE combinations at different positions in the image. In this thesis this has only briefly been investigated and therefore not included in the final results.

4.1.4 Intensity Differences

Objects in the real world can by the human eyes easily be distinguished from each other depending on how close they appear and by the color intensity they possess. Separating objects in an image based on intensity differences between

pixels is commonly being used to extract information. For instance, if there is a distinguishable difference in intensity between two adjacent pixels there is a high probability that these two pixels belong to different objects in the image. Calculating the difference in intensity of an image can for instance be done by calculating the *Gradient Image* (see Figure 4.4).

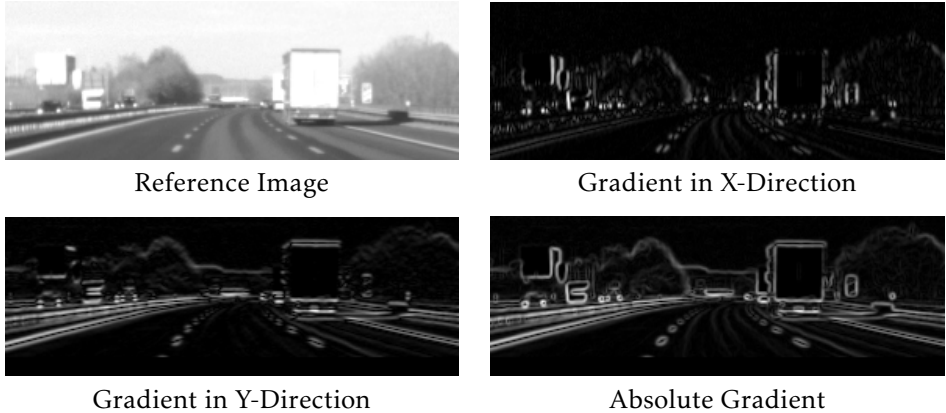


Figure 4.4: Gradients extracted from reference image.

In order to include intensity differences between pixels in the energy function the following likelihood function is applied:

$$L_{int} = \begin{cases} 1, & \text{if } \Delta_I(p, q) < I_\mu \\ \exp\left(-\frac{(\Delta_I(p, q) - I_\mu)}{I_s}\right), & \text{otherwise} \end{cases} \quad (4.5)$$

where $\Delta_I(p, q)$ is the absolute value of the difference in image intensity between the two pixels in question, I_μ is the median of the pixel intensity difference over the ROI and I_s is the average deviation from I_μ . The distribution applied has the appearance of a non normalized Laplace distribution.

4.2 Initialization of Graph Cuts

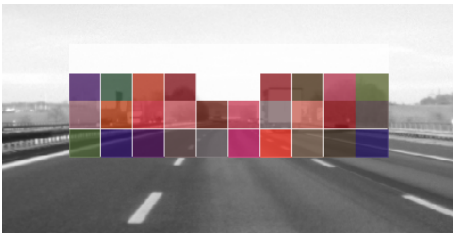
One of the disadvantages of using graph cuts and pairwise MRF or CRFs for image segmentation in general is that inference methods are dependent on initialization. Initialization affects run time and can, if poorly chosen, give results found as unlikely. Because of this there is a need for a suitable initialization which approximates the final result, giving a lower energy in the MRF/ from start. In this thesis a rectangular grid and the results of a super pixel segmentation algorithm has been investigated as initialization methods. For initializing the plane parameters, a RANSAC fit to an initial stereo result (e.g. SGM) is performed.

4.2.1 Rectangular Grid

One initialization method is to approximate the world with a certain number of equally large rectangular elements located at a distance extracted using a separate stereo algorithm (e.g. SGM). Obviously, by using this method shapes of objects are not followed in any way. This initialization method can give bad results if there is a possibility for the inference methods to keep the rectangular shapes without having an corresponding object in the image. Figure 4.5 (a) illustrates a rectangular grid initialization.

4.2.2 Superpixel Segmentation (gSLIC)

A more sophisticated initialization method is to use the results from a superpixel segmentation algorithm. A superpixel segmentation algorithm divides the entire set of pixels in the image to a subsets of connected pixels, called superpixels. Forming compact and uniform superpixels can be done using GSLIC [22], which is a parallel implementation of the SLIC superpixel segmentation [1]. SLIC initializes cluster centres and moves the centres to the lowest gradient position within a small neighborhood. The clusters are then iteratively grouped into superpixels based on intensity values and spatial proximity. The algorithm is considered to provide real time performance and to provide legitimate initialization to inference methods. Figure 4.5 (b) illustrates this initialization method.



(a) Rectangular Grid Initialization.



(b) gSLIC Initialization.

Figure 4.5: Initialization methods. PSEs which are located where no disparity values are available is assigned a white color.

4.2.3 Position Dependent Initialization

In the case of long-range road scenes the likelihood of finding smaller objects increases with the distance. This assumption gives that the further away from the camera an objects is located it will, if found, be represented by smaller (in the two dimensional image plane) PSEs. To be able to find objects at a large distance the initialization might therefore require a large amount of PSEs. The run time is however increased by considering more PSEs. To be able to have an initialization which has smaller PSEs at larger distances and which does not increase the number of initial PSEs significantly this thesis considers an initialization which have

larger PSEs in the outer part of the ROI and smaller PSEs in the inner part of the ROI. Figure 4.6 (a) and 4.6 (b) illustrates this.

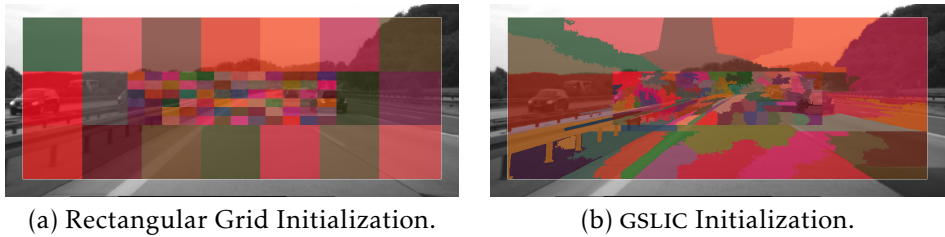


Figure 4.6: Position dependent initialization methods.

4.3 Restriction of Expansion Areas

In inference methods based on standard graph cuts every vertex has the possibility to be assigned any of the estimated PSEs in order to minimize the defined energy. This is however not efficient since the possibility for a PSEs to be present in the entire ROI is rather unlikely, especially for a long-range road scene application. For instance if there would be an occluded vehicle at a large distance the PSE representing the vehicle would be very small. Because of this there is a need to discard areas of the image while performing graph cuts on certain PSEs. Besides the obvious advantage of a faster inference method, restricting the area where it is possible for planar scene elements to expand to also gives the advantage of reducing the likelihood of finding disjoint PSEs, i.e. PSEs which are divided into several parts of the image. Obtaining a solution with disjoint PSEs can occur when the scene contains two objects at the same distance from the camera located at different positions in the image. The restrictions of expansion areas are reasonable restrictions of the solution space. The PSEs are still allowed to grow beyond the initial restrictions, since they are adapted in each iteration.

In this thesis the following expansion areas are investigated:

- Neighboring PSEs.
- Rectangular areas around the center point of the considered PSE.
- Ground and sky allowed to expand to their respective side of the horizon.
- PSE representing generic objects restricted to areas where they are above ground.

There is also the possibility to use any combination of these restrictions (restricting the expansion to the intersecting area between them). In addition to the the restrictions mentioned PSEs can be restricted to expand to areas where their respective depth value is within a valid depth interval. This interval is defined

from Z_{min} to Z_{max} found in table 4.3. By using this restriction run time can be decreased since vertices which have a probability of zero and should therefore never appear in the solution are never checked. For instance, a PSE representing a crash barrier is almost parallel to the viewing direction and does therefore contain depth values which are larger than the allowed maximum distance. Initialization methods provided in this thesis do not consider the valid depth interval when initializing and can for that reason initialize the graph with values which should have a probability of zero. If this restriction is not applied there might exist pairwise potentials which do not fulfill the condition of submodularity as defined in (2.8). Restricting PSEs to certain areas is resulting in true hard constraints which forbids PSEs to be present in certain parts of the image. This is something which otherwise can not be guaranteed by assigning a very low probability, i.e. assigning a very high cost in the energy function.

Restriction of expansion areas is very important for this thesis since it reduces computational time significantly and improves the results.

4.3.1 Expansion to Neighboring Planar Scene Elements

One restricted area for each PSE to expand to is to the neighboring PSEs area. This restricts the PSEs expansion area relatively much (depending on the number of initial planar scene elements), giving less vertices to consider during inference. The downside of this method is that the expansion area can vary depending on the shape of the neighboring scene elements. For instance if a certain planar scene element is a neighbor of ground or sky there is very large areas which it can expand to comparing to a scene element which is not a neighbor of sky or ground.

4.3.2 Expansion to Rectangular Area

In order to expand to a area not depending on the shape of the neighboring PSEs this thesis considers an approach which allows expansion to a rectangular area around the considered PSE. This method estimates the center point of each planar scene elements and allows expansion to a rectangular window (with a user defined width and size, which is a factor of the smallest bounding box containing the entire PSE considered) around the estimated center point. This restriction is illustrated in Figure 4.7.

4.3.3 Expansion limited by Horizon

In order to decrease run time while expanding PSEs representing ground and sky one can discard the parts of the image which is not valid, i.e ground should not be able to expand to the area above the horizon and sky should not be able to expand to the area below the horizon. This assumption is already found in the model prior, however the model prior is not restricting the expansion. This gives that multiple checks in the expansion of ground and sky are unnecessary. The

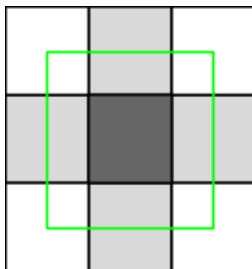


Figure 4.7: *Illustration of restriction of expansion areas. If rectangular expansion is applied expansion for the center PSE is limited to expand within the defined green rectangle in one iteration.*

impact on the final results of this restricted expansion is thereby not clearly noticeable, however the run time is decreased. The run time effect of this restriction is increasing with the size of the ROI.

4.3.4 Expansion limited by ground

One of the assumptions made for this labeling task is that objects should not be able to be found within the ground. Objects should therefore not be able to expand to areas which are located below the ground plane. This situation is modeled as a approximate hard constraint in the pair potentials but can still occur in the solution. To include this as a real hard constraint this thesis considers to restrict PSEs representing objects to have a depth which is larger than ground at all pixel positions below the estimated horizon.

5

Results

The previous chapters have provided the necessary background information and implementation details regarding this labeling task. This chapter will present and evaluate the results obtained from MRF/CRF approaches. It will present results with slightly different parameter settings to show the impacts of a more sophisticated model. The results are evaluated and compared to other state-of-the-art segmentation algorithms. In the example images below, ground is given a green color and sky is given a blue color. Other PSEs are assigned a random color. In the results a modified alpha-expansion graph cut method which incorporates fixed label costs is used for inference [6][18][4][8].

5.1 MRF Results

The MRF results are presented below and are included in the evaluation in Section 5.3. The results are not considered as good as the results of the CRF approach with more sophisticated binary potentials but yet included in order to show the improvement when considering richer interactions between vertices and the importance of restriction areas for PSEs. As can be seen in Figure 5.1 (left) there exists false object detections on the ground plane. This is because a rather simple method is used for computing the applied likelihoods (see 3.3), which is sensitive to illumination changes between the cameras. In other words, the brightness constancy assumption is violated. A simple MRF with no restrictions is not sufficient to eliminate the impact of illumination changes between the cameras.

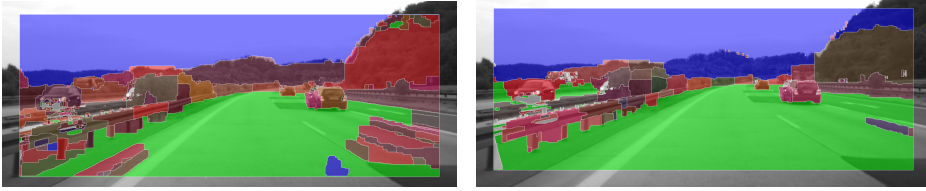


Figure 5.1: Labeling results of a large ROI using MRFs with a Potts model assigned. No restriction of expansion areas for PSEs (left) and all restriction applied (right)

5.2 CRF Results

The resulting images from a CRF approach is provided below. Comparing to the MRF results in Figure 5.1 the main difference is that there is no false positive objects near the vehicle and the estimation of ground is improved at larger distances. A problem regarding both the MRF and CRF results is the time required to obtain results. The bottleneck is the graph based inference method.

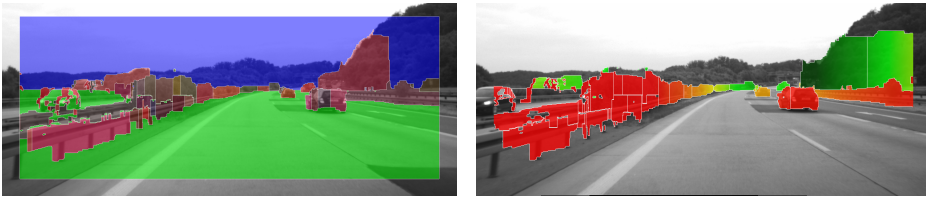


Figure 5.2: Labeling results of a large ROI using CRF. The left image shows found PSEs (ground green, sky blue and other objects are assigned a random color). The right image shows distances assigned to found objects. Red indicates a close distance and green indicates a far distance.

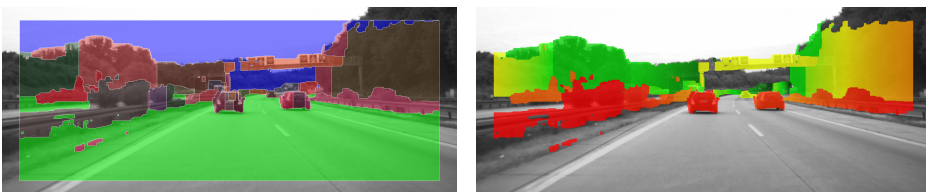


Figure 5.3: Labeling results of a large ROI using CRF. The left image shows found PSEs (ground green, sky blue and other objects are assigned a random color). The right image shows distances assigned to found objects. Red indicates a close distance and green indicates a far distance.

5.3 Evaluation

The algorithm based on graph cuts presented in this thesis are compared to state-of-the-art segmentation methods. The evaluation performed is mainly a comparison between the predicted results and a data set representing ground truth values. The evaluation is designed in a specific way in order to compare the results with other methods, including results from the original stixel world and an improved stixel world [7]. The improved stixel world uses object-level priors in order to improve the results. At far distances an appearance based vehicle detector is used to improve the stixel results and thereby performance. The improved stixel world is in the tables giving performance measures notated as *StixFix* and the original stixel world as *original stixels*. The results are also compared to state-of-the-art superpixel segmentation methods, namely SLIC [1] and the graph-based image segmentation method (GBIS) in [11]. The methods have been evaluated on a large ROI, as seen in Figure 5.1 and Figure 5.2. A small ROI (half the size of the large ROI) is also evaluated considering freespace estimation.

5.3.1 Ground Truth Dataset

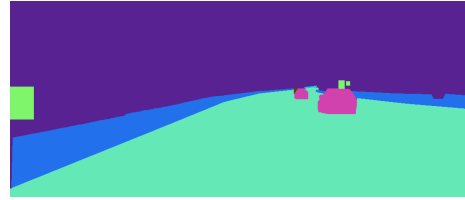
The data set used for evaluation is a set of 2000 manually labeled frames representing a German highway. The data set is an extension of the data set used for evaluation in [7]. The highway contains various vehicles at different distances. Every tenth frame contains a complete and pixel-accurate labeling of the road environment. In the rest of the frames, only vehicles are labeled using polygons. Every tenth frame also contains more classes labeled, such as crash barriers and the road. Figure 5.4 (a) illustrates more accurately adjusted bounding boxes for vehicles. Figure 5.4 (b) shows ground truth labels for a complete and pixel-accurate labeling. Figure 5.4 (c) shows ground truth for all objects for the same frame as in Figure 5.4 (b). Note that the generated ground truth data set is not exactly following the borders of all objects for each frame, however it is considered accurate enough to provide a legitimate performance measure for pixel-wise labeling of road scenes. The data contains vehicles and occluded vehicles at very large distances, as can be found in Figure 5.4 (d). The difference performance measures applied when comparing results and the generated ground truth data set are further explained below.

5.3.2 Segmentation Accuracy and Object Detection Rate

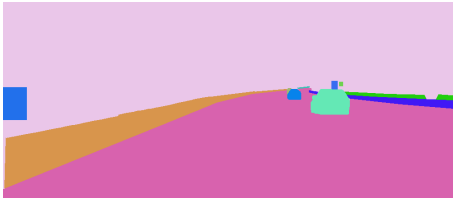
In order to compare the performance of the methods presented in this thesis, a measure of segmentation accuracy has to be defined. Segmentation accuracy is found by comparing the estimated objects with ground truth values. The performance measure is according to the evaluation in [7]. A intersection-over-union (IU) for each object is calculated individually and averaged over all objects of the same class. The more traditional PASCAL VOC IU [10] is not used because objects which are located further away (and is thereby containing fewer pixels) should have an equal impact on the performance measure as objects located at a closer



(a) Ground truth values for vehicles.



(b) Ground truth classes.



(c) Ground truth objects.



(d) Enlarged area of ground truth objects.

Figure 5.4: *Ground Truth Images. As can be seen, the data set contains rather difficult objects to segment accurately.*

distance.

Thresholding the IU with 0.5 gives a detection rate, i.e if the estimated object overlaps more than 50 percent with the ground truth object it is counted as detected. The segmentation accuracy presented is for the class non-occluded vehicles since this is considered as the most important class for highway scenarios. The class vehicle includes both cars and trucks in this evaluation.

The detection rate (MRF Figure 5.5 and CRF Figure 5.6) clearly shows the potential of this method. Especially the IU over the number of superpixels is a lot better than the rest. This is considered good since the actual size of the vehicles are interpreted more accurately and the confidence of actually finding vehicles are higher. With the simple Potts model it is possible to achieve a detection rate of over 0.8 which is comparable to the other methods. The advantage of the methods presented in this thesis is that fewer number of superpixels is describing the vehicles. The difference between a MRF and a CRF is not really visible (since false positives do not contribute in this performance measure).

5.3.3 Freespace Estimation

Freespace is in this scenario the part of the image which is not occupied by objects or sky, i.e ground not occupied by objects. Evaluating freespace detection is of interest since it gives an indication of how well objects are detected independent on the appearance of the detected objects and if they are incorrectly divided into more than one PSE. A freespace estimation is also of interest since it indicates how well the methods finds areas where the autonomous vehicle should be able to drive without colliding with other obstacles and the impact of false positives detections. In order to achieve a freespace evaluation which is comparable to

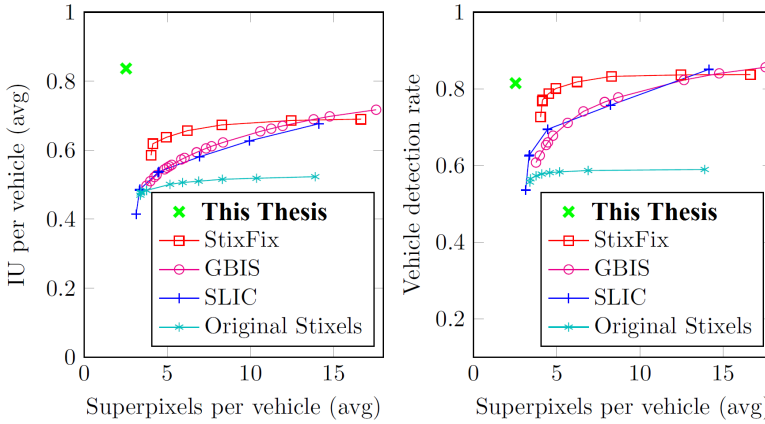


Figure 5.5: A MRF approach compared to other state-of-the-art segmentation methods in terms of segmentation accuracy over segmentation complexity for the class vehicle. The latter is expressed by the average number of superpixels per object. Accuracy is compared by providing the upper limits for any system based on these superpixels using the average intersection-over-union (IU) per object (left) and the detection rate (right). The markers stands for different parameter sets and the solid lines connects the best performance.

other methods, the vertical limit of the freespace in front of the autonomous vehicle is marked for each image column in every tenth frame. The performance measure is the detection rate of the freespace limit. For each column in the ROI, a column count as detected if the deviation of the estimated limit to ground truth is within a defined threshold (given in pixels).

As can be seen in Figure 5.7 the detection rate of freespace in the MRF approach is very low (indicating many false object detections) in comparison to the original stixel world and to the improved stixel world. This is because without any world constraints incorporated in the model unlikely PSEs *cannot* be discarded. This can clearly be seen in Figure 5.1 where PSEs are present in the bottom part of the image where it in fact should be ground (because of the violation of the brightness constancy assumption). To prove the importance of restricting expansion areas for PSEs the freespace detection rate for a MRF is given twice in Figure 5.7. The poorest performance have no restrictions applied and the best performance have all restrictions applied. In the CRF results, where all restrictions and the limited scene model are applied the freespace estimation is very close to the original stixel world but what is desired is of course a method which outperforms the improved stixel world.

Evaluation of Small ROI

In the real world many object occurrences are impossible. In this thesis these occurrences are modeled by assigning a very high cost. This is however not al-

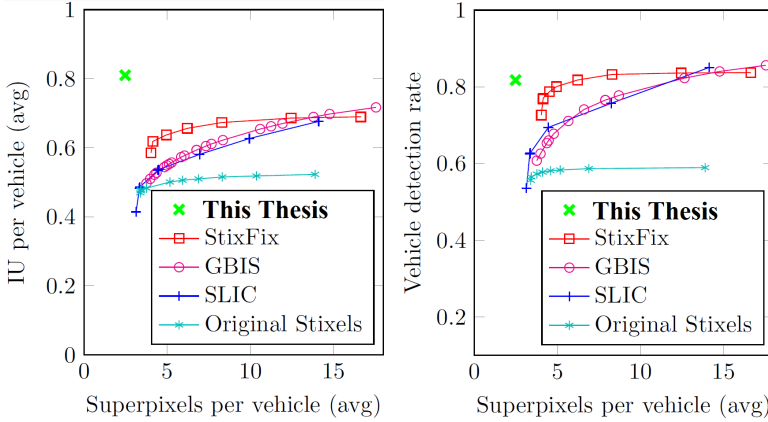


Figure 5.6: A CRF approach compared to other state-of-the-art segmentation methods in terms of segmentation accuracy over segmentation complexity for the class vehicle. The latter is expressed by the average number of superpixels per object. Accuracy is compared by providing the upper limits for any system based on these superpixels using the average intersection-over-union (IU) per object (left) and the detection rate (right). The markers stands for different parameter sets and the solid lines connects the best performance.

ways sufficient since the value assigned to represent impossible is in fact only an approximation. This problem shows in the freespace evaluation. As an example, crash barriers in the ground truth data set do not follow the crash barriers in reality completely. Crash barriers are instead estimated to have a rectangular shape (see Figure 5.4). This is a disadvantage for the methods presented in this thesis since crash barriers are sometimes estimated more accurately than the ground truth data set. This gives that the freespace evaluation gives a lower detection rate when in fact it should be higher. As can be seen in Figure 5.8, the PSE is accurately following the borders of the crash barrier.

To compare with the results given in [7], a rather large ROI is considered in the case of obstacle detection for long-range road scenes. One disadvantage of this is that there is not always valid disparity values at the edges of the ROI, causing the method not to find any obstacles at this area. This scenario can be seen in Figure 5.8. This problem does however also occur for stixels since the stixel world is also derived from disparity values. In the case of the improved stixel world they use a hough transform which can in some cases provide a better estimation of the crash barrier in this area and thereby finding a better detection rate for ground. In this thesis focus has been on creating an application on long-range road scenes. Because of this a detection rate for ground of a smaller ROI representing long-range road scenes is given.

In Figure 5.9 the strengths and potential of this method is visible. The detection rate is outperforming the improved stixel world for long-range road scenes.

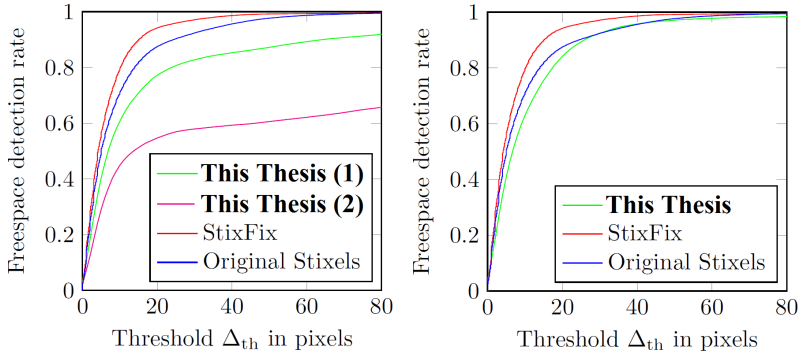


Figure 5.7: Freespace detection rate of a large ROI. Original stixels and stix-Fix are compared to the MRF approach (left) with the potts model assigned and a CRF approach (right) which has a richer interaction between vertices. The left graph includes two performance measures. The one performing the worst is a MRF which allows PSEs to expand to areas below ground.

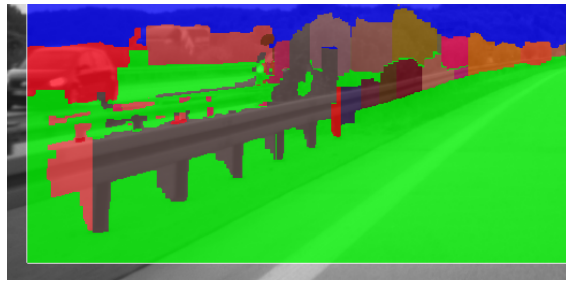


Figure 5.8: Bottom left area of large ROI.

Evaluation of Sparse Stereo

In the case of the stixel world, a dense stereo matching is assumed. In the presented method there is however no assumption regarding this. This gives that it is of interest to know if the initial stereo method is crucial. Because of this a performance measures using a correlation based stereo method instead of SGM is of interest. The performance measures obtained shows that there is no significant difference between the results (see Figure 5.10). The reason for this is that the obtain disparities are only used for initialization. Having a method not dependent on dense stereo matching is however desirable.

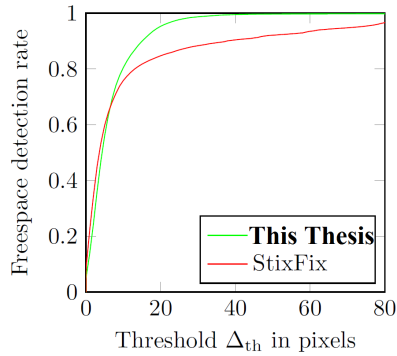


Figure 5.9: Freespace detection rate of small ROI representing long-range road scenes.

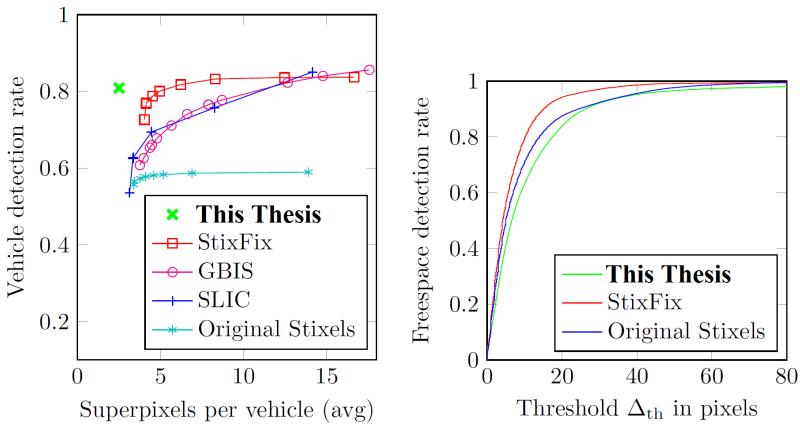


Figure 5.10: Vehicle detection rate using sparse stereo matching (left) and freespace detection rate of large ROI using sparse stereo matching (right). Note that it is only this work which uses sparse stereo matching, other methods in the figure above is obtained with dense stereo matching. Comparing the detection rate above with the vehicle detection rate with the vehicle detection rate of this paper in Figure 5.6 no significant difference is found.

6

Discussion and Future Work

The goal of this thesis is to perform long range obstacle detection. This is approached by performing pixel-wise labeling of gray-scaled images representing long-range road scenes. The overall method for this thesis includes estimating 3D planes (PSEs) from stereo image pairs, introduction of geometric constraints and restriction of the solution space. The introductions of geometric constraints in the labeling task is the main advantage of this method. The estimation of PSEs gives data likelihoods for each pixel to belong to each of the estimated PSEs.

The data likelihoods and a simple Potts model (2.5) applied are not strong enough to provide a solution with accurate results. This is mainly due to the impact of illumination changes between the two cameras and the image noise. Because of this geometric constraints are introduced which facilitate the graph based optimization method to make correct assignments of pixels. Restriction of expansion areas for PSEs are introduced to reduce the solution space. The restrictions are of high importance since they reduce the run time of the optimization and removes unwanted solutions. The solution space is however still very large with the restrictions which gives computationally heavy inference. Therefore, a dynamic programming approach is briefly investigated in this thesis (see Appendix A).

The methods provided in this thesis show that it is possible to obtain spatial perception which outperforms state-of-the-art algorithms. Benefits of this solution compared to the stixel world [3] is that it is possible to obtain slanted surfaces and the fact that the PSEs are directly extracted from the stereo image pairs, not the SGM results. This chapter will discuss the overall methods presented in this thesis, give a glimpse of future work in the area of the thesis and briefly visit the ethical and social aspects of the thesis.

6.1 Method

The methods presented in this thesis show good results and that the likelihoods provided as input data can be used to segment images representing road scenes and detect objects at far distances. The major downside is the time required to obtain results. In the current stage of the methods there is no possibility to include them in the autonomous vehicle. Obtaining results for a small ROI takes roughly one second. In other words, the time required to find objects is more or less the same as it takes for the vehicle to travel the distance to where the object is located. The bottleneck is the inference method based on graph cuts.

6.2 Results

The results which are obtained from a MRF approach that only applies the unary potentials and the potts model are outperformed by the original stixels [3] and stixFix [7] when it comes to detection rate of ground. The vehicle detection rate is however quite similar, clearly showing the strength of the provided likelihoods. When more information is included in the smoothness term results are improved, the detection rate for ground is increased and false object detections are reduced significantly. The results show that when considering long-range road scenes the detection rate for ground is significantly higher than the improved stixel world with object-level priors, clearly showing that the unary potentials provided can increase the range to when objects are first detected. The results also show that the stereo method used for initialization is not critical.

6.3 Future Work

In order to have an algorithm which can be used for long-range obstacle detection in the autonomous vehicle, a faster inference method which gives good results and which works in real time must be provided (e.g. a DP approach). The methods presented in this thesis are too time consuming in order to be implemented in the autonomous vehicle. The methods presented in this thesis does however show that with the input data provided and with a limited scene model applied results can be obtained which outperforms other state-of-the-art segmentation algorithms. It should therefore be possible to apply the same input data and a similar world model in order to obtain a faster method based on DP which shows similar results. It should also be possible to reduce the run time of the current graph based method by moving from the current pixel based labeling to super-pixels or patches as the basic elements in a graph representation. This would reduce the number of nodes in the graph and thereby reduce run time.

The methods presented in this thesis only considers the image frame currently being labeled when generating the unary potentials. For long-ranged road scenes this likelihood is not considered strong due to the low SNR at larger distances.

It is however possible to include multiple frames when estimating the PSE likelihoods. This would in theory make the likelihoods, which now are rather weak, stronger and thereby giving the advantage to label obstacles at larger distances more accurate. It is also possible to use a more robust matching cost for computing the likelihoods. Having stronger unary potentials is beneficial considering the complexity of the smoothness term. The stronger and more accurate the unary potentials are the less complex the smoothness term needs to be in order to segment the image accurately.

In the CRF approach many potentials can be added and adjusted. In this thesis there has not been any major focus on finding the optimal combination of the included costs. For instance the potentials given by a distribution (see Table 4.3) should in theory improve the results. Potentials depending on PSEs relation to the estimated horizon can also be added in order to improve the results. In the current state of the initialization hard constraints are not considered. This gives that there is a possibility to obtain better initializations which can alter the solution space. Hard constraints are also not considered when estimating parameters. Incorporating the hard constraints while estimating the parameters would in theory improve the method, i.e unlikely combinations of PSEs can be removed earlier while inferring.

Another way to improve results would be to introduce a cost based on learning. This would in theory provide probabilities of whereabouts of objects. For instance, crash barriers on highways often have a typical appearance and size and it would of course make sense to use this information in order to obtain better results. In this thesis learning has not been investigated.

6.4 Ethical and Societal Aspects

The main advantage of autonomous driving is the possibility to reduce the human factor in daily traffic and thereby saving lives. In 1997, the Swedish government set the goal that no one should be killed or seriously injured in traffic. This is known as the *Zero Vision* (Swedish Nollvisionen). I believe that in order to reach this goal the autonomous vehicle must be introduced and become a normal part of our lives. Besides saving lives, the autonomous vehicle can reduce traffic jams and driving time. This is of high interest since it will lead up to less pollution and thereby a greener planet. To conclude, for society in general the autonomous vehicle would contribute to a safer environment.

One of the major discussions regarding introducing the autonomous vehicle and robotics in general is the question of who carries the responsibility when the vehicle or robot causes damage to a person or in the worst case kills a person. This is a very difficult discussion since it involves many aspects. First of all, is it the driver who should be responsible? Should the company selling the vehicle be responsible? What happens if two autonomous vehicles collide, which car is then responsible? Today there are regulations about responsibility for the driver,

but in the future there might not be a driver but a vehicle which makes all the decisions. This would mean that the insurances and laws we have today must be reformed. This discussions is on going and will need a lot more attention before the autonomous vehicle can be introduced and available to the public.

6.5 Conclusions

The results shows that using the likelihoods for a pixel to correspond to a finite set of PSEs it is possible to perform obstacle detection and image segmentation accurately. The obstacle detection is also working at a larger distance where it outperforms the improved stixel world with object-level priors in freespace detection rate. The run-time of the methods presented in this thesis are however to high for a practical use.

The results also show that it is possible to introduce a limited scene model which reduces the solution space. This is of high interest since it indicates that a limited scene model working together with an efficient DP approach should be possible to obtain. It should thereby be possible to obtain good results with different solution methods.

To reduce run-time, an efficient DP approach must be implemented which is not as time consuming as a graph based solution or the DP approach (see Appendix A) presented in this thesis. A DP approach would be possible since many of the limitations applied in the graph based methods can also be applied for a DP approach. For the nearest future it would be exciting to see if a DP approach can be fast enough and thereby be implemented in the autonomous vehicle.

Appendix

A

Labeling with Dynamic Programming

Using generic inference algorithms on MRFs when solving the multi-class labeling problem can be a problem when considering time complexity and accuracy of the model representing the scene. Therefore, a dynamic programming (DP) approach is introduced as an alternative method for solving the multi-class labeling problem. In general, DP is a method where a solution to a problem is obtained by identifying a collection of subproblems and solving them one by one, starting with the smallest, and using the results from these to find the answers to larger subproblems. One advantage of using a DP approach as used in [12] is that it will always give an optimal labeling of the scene. This means that it is not dependent on initialization as in the case of a graph-cut based implementation.

Any DP approach is usually constricted to a one dimensional or low tree width structure. In computer vision applications, such as the multi-class labeling problem, this is a problem since most domains are in two dimensions or higher [12]. Because of this a DP approach must have additional constraints on the scene which reduces the dimensions of the solution space. Examples of this can be found in [12] and [25]. The main goal of a DP approach for the labeling task is to minimize the function:

$$E = \sum_P D_p(l_p) + \lambda \sum_{(p,q) \in \mathcal{N}} V(l_p, l_q) \quad (\text{A.1})$$

This function corresponds to the MRF energy defined in equation 2.4, containing a data cost and pairwise potentials for increasing the robustness of the labeling when the data cost is insufficient [27]. The problem regarding how to design pairwise potentials remains. A good trade off between the data which is to be included must be found.

A.1 Tiered Scene Labeling

The method presented in [12] enforces a *tiered* structure to decrease the solution space. A tiered structure means that the image is expected to have larger regions which holds certain relations to each other. In [12] the image is divided by two horizontal curves creating a *top*, *middle* and *bottom* region of the image (see Figure A.1). The middle region is further vertically divided into subregions where the subregions is assigned a label from a finite set of labels.

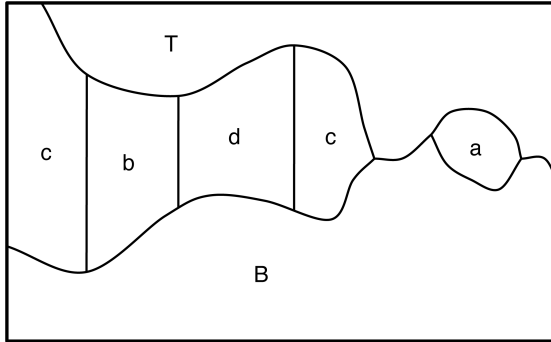


Figure A.1: Illustration of Tiered labeling. [12]

In [25] the scene is restricted to contain five regions with a given relation to each other. The model enforces the regions (*center*, *top*, *bottom*, *left* and *right*) to maintain at the given relation to each other by assigning a penalty of infinity for certain labeling occurrences. For instance if a pixel is labeled *bottom*, it should not be able to have a neighbor directly beneath it to be labeled as *top*.

Because of the similarities of the three-tiered structure scene with long-range road scenes the algorithm presented in [12] is used as a baseline for a implementation with a DP approach for solving the multi-class labeling problem.

The three-tiered DP method always reaches global optimum, which is demanding and requires computational time. For long-range road scene scenarios, and especially if the focus is on freeways, there is a need for fast computations. If computation is considered too slow, this method must be discarded. Another problem with the three-tiered scene labeling presented in [12] is that the middle region is divided vertically, only allowing one object to be present in each column of the middle region. This is not desirable, scenarios where one object is located further away than another but both are present in the same column of the image can and will occur for long-range road scenes. For instance if a truck is in front of a car, the algorithm will not find the truck and the car as two individual objects.

A.2 Fast Tiered Labeling

In addition to the disadvantages of the three-tiered scene labeling, it is quadratically dependent on the number of labels within the middle tier. This is not

desirable for a road scene application since the number of labels depend on the number of objects currently present in the scene. In a scenario where multiple objects can be present and where computational time is of high interest there exists a desire for a fast algorithm. A simplified version of the tiered structure in [12] is suggested in [27]. The method in [27] provides a significantly faster algorithm which approximates the exact labeling of an image. It also provides the possibility for a middle region to be horizontally divided between objects, it is thereby not restricted to have one label per column in the middle region. However, the cost of obtaining this is the loss of certainty of reaching global optimum.

The method in [27] makes three simplifications. First the cost of multiple labels within a single tier, k , is aggregated into a single cost function, D_k . For each pixel p , the cost function is defined as:

$$D_k(p) = \min_{l \in L_k} D_p(l_p) \quad (\text{A.2})$$

where L_k is the labels within tier k and D_p is the data cost for label l_p at pixel p .

The second simplification splits the entire labeling problem into binary labeling problems. Starting from the bottom tier, it separates the bottom tier from all other tiers and then continue to the next. This creates $K - 1$ binary problems, given that the number of tiers is K (see Figure A.2).

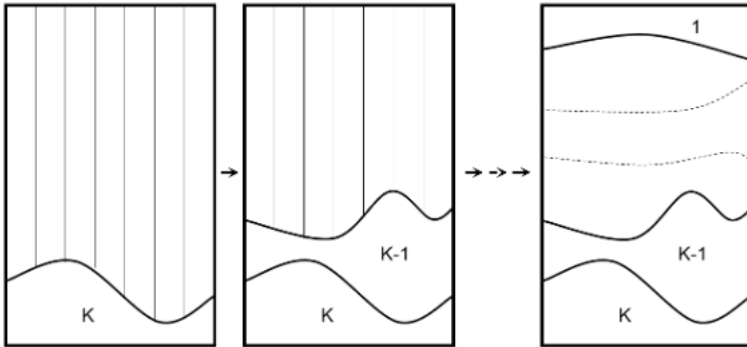


Figure A.2: Illustration of binary labeling. [27]

The third simplification made restricts the pairwise potentials. Vertical pair potentials (potentials between pixels in the same column) can be arbitrary chosen but horizontal pair potentials (potentials between pixels in the same row) is given the form of the potts model as can be found in equation 2.5.

The method also introduce the possibility to use the number of extrema of a path to quantify its topological smoothness. In other words, it uses the number of peaks and valleys of a curve separating two tiers. This is something that cannot be done using traditional MRFs [27].

This method is however not entirely appropriate for labeling long-range road scenes based on the likelihood data provided for this labeling task. The first simplification gives no possibilities to exploit relations between the estimated planar

scene elements found in the scene. The strengths of the estimated planar scene elements are thereby discarded. Because of this there is a need for an alternative DP approach which is faster than the three-tiered algorithm and uses the planar scene elements provided.

A.3 Tiered Scene Labeling Results

The DP approach presented in [12] is showing promising results for image segmentation (see Figure A.3 (a)). The fact that it is quadratically dependent on the number of elements within the middle tier is however causing a problem, computational time for an application considering labeling objects in a road scene scenario is needs to be low. The limited scene model is also a problem in a road scene scenario. In some cases objects are found where lane markings are located. This can be seen in Figure A.3 (b). Because of this and since the MRF and CRF approaches showed more promising results the DP approach is included as an appendix in this thesis and remains a topic of future research.

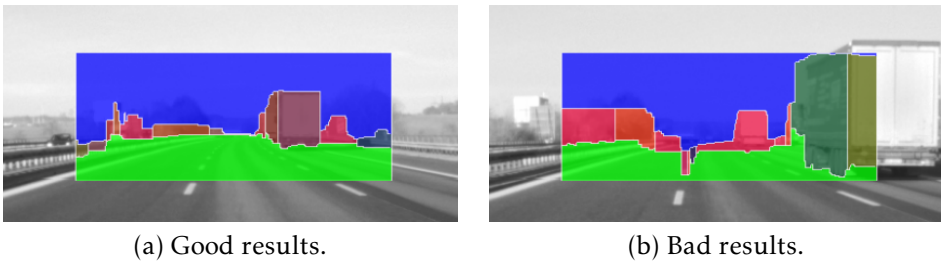


Figure A.3: Labeling results of a smaller ROI using a DP approach with 15 initial elements.

Bibliography

- [1] R. Achanta, A. Shaji, K. Smith, A. Lucchi, P. Fua, and S. Ssstrunk. SLIC Superpixels Compared to State-of-the-Art Superpixel Methods. *IEEE TPAMI*, November 2012. Cited on pages 32 and 39.
- [2] B. Andres, T. Beier, and H. Kappes. OpenGM: A C++ Library for Discrete Graphical Models. *ArXiv e-prints 2012*, June 2012. Cited on page 17.
- [3] H. Badino, U. Franke, and D. Pfeiffer. The Stixel World - A Compact Medium Level Representation of the 3D-World. *31st DAGM Symposium on Pattern Recognition*, September 2009. Cited on pages 2, 4, 45, and 46.
- [4] Y. Boykov and V. Kolmogorov. An Experimental Comparison of Min-Cut/Max-Flow Algorithms for Energy Minimization in Vision. *IEEE TPAMI*, September 2004. Cited on page 37.
- [5] Y. Boykov, O. Veksler, and R. Zabih. Fast approximate energy minimization via graph cuts. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, November 2001. Cited on pages 16 and 17.
- [6] Y. Boykov, O. Veksler, and R. Zabih. Efficient Approximate Energy Minimization via Graph Cuts. *IEEE TPAMI*, November 2001. Cited on page 37.
- [7] M. Cordts, L. Schneider, U. Franke, and S. Roth. Object-level Priors for Stixel Generation. *GCPR*, November 2014. Cited on pages 4, 39, 42, and 46.
- [8] A. Delong, A. Osokin, H. N. Isack, and Y. Boykov. Fast Approximate Energy Minimization with Label Costs. *IEEE CVPR*, June 2010. Cited on pages 22 and 37.
- [9] M. Enzweiler, M. Hummel, D. Pfeiffer, and U. Franke. Efficient Stixel-Based Object Recognition. *IEEE Intelligent Vehicles Symposium*, 2012. Cited on page 4.
- [10] M. Everingham, L. V. Gool, C. K. I. Williams, J. Winn, and A. Zisserman. The Pascal Visual Object Classes (VOC) Challenge. *International Journal of Computer Vision* 88, June 2010. Cited on page 39.

- [11] P. F. Felzenszwalb and D. P. Huttenlocher. Efficient graph-based image segmentation. *International Journal of Computer Vision* 59, September 2004. Cited on page 39.
- [12] P. F. Felzenszwalb and O. Veksler. Tiered scene labeling with dynamic programming. *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, June 2010. Cited on pages 51, 52, 53, and 54.
- [13] S. K. Gehrig, F. Eberli, and T. Meyer. A Real-Time Low-Power Stereo Vision Engine Using Semi-Global Matching. *Proceedings of the 7th International Conference on Computer Vision Systems: Computer Vision Systems*, October 2009. Cited on pages 2 and 10.
- [14] J.M. Hammersley and P. Clifford. Markov fields on finite graphs and lattices. *Technical Report*, 1971. Cited on page 21.
- [15] H. P. Hirschmüller, R. Innocent, and J. M. Garibaldi. Real-time correlation-based stereo vision with reduced border errors. *International Journal of Computer Vision*, April-June 2002. Cited on page 10.
- [16] H. Hirshmüller. Accurate and efficient stereo processing by semi-global matching and mutual information. *IEEE Conferance on Computer Vision and Pattern Recognition (CVPR)*, June 2005. Cited on pages 2, 10, and 11.
- [17] V. Kolmogorov and C. Rother. Minimizing non-submodular functions with graph cuts – a review. *IEEE PAMI*, June 2007. Cited on page 17.
- [18] V. Kolmogrov and R. Zabini. What Energy Functions can be Minimized via Graph Cuts? *IEEE TPAMI*, February 2004. Cited on page 37.
- [19] S. Z. Li. *Markov Random Field Modeling in Image Analysis. Advances in Pattern Recognition*. Springer, 2010. Cited on page 21.
- [20] D. Pfeiffer and U. Franke. Towards a Global Optimal Multi-Layer Stixel Representation of Dense 3D Data. *Proceedings of the British Machine Vision Conference 2011*, August 2011. Cited on pages 3 and 24.
- [21] R. Szeliski. *Computer Vision Algorithms and Applications*. Springer, 2011. Cited on pages 9, 10, 13, 15, 18, and 21.
- [22] C. Y. Ren and I. Reid. gSLIC : a real-time implementation of SLIC super-pixel segmentation. *Technical Report, University of Oxford, Department of Engineering Science*, June 2011. Cited on page 32.
- [23] C. Rother, S. Kumar, V. Kolmogorov, and A. Blake. Digital Tapestry. *IEEE CVPR*, June 2005. Cited on page 17.
- [24] D. Scharstein and R. Szeliski. A taxonomy and evaluation of dense two-frame stereo correspondence algorithms. *International Journal of Computer Vision*, April 2002. Cited on page 10.

-
- [25] O. Veksler. Fast dynamic programming for labeling problems with ordering constraints. *IEEE Conference on Computer Vision and Pattern Recognition*, June 2012. Cited on pages 51 and 52.
- [26] C. Wang, N. Komodakis, and N. Paragios. Markov Random Field modeling, inference & learning in computer vision & image understanding: A survey. *Computer Vision and Image Understanding*, November 2013. Cited on pages 9, 11, 12, 13, 15, 16, 18, and 21.
- [27] Y. Zheng, S. Gu, and C. Tomasi. Fast Tiered Labeling with Topological Priors. *The 12th European Conference on Computer Vision*, October 2012. Cited on pages 51 and 53.
- [28] J. Ziegler, P. Bender, M. Schreiber, H. Lategahn, T. Strauss, C. Stiller, T. Dang, U. Franke, N. Appenrodt, C. G. Keller, E. Kaus, R. G. Herrtwich, C. Rabe, D. Pfeiffer, F. Lindner, F. Stein, F. Erbs, M.ENZweiler, C. Knoppel, J. Hipp, M. Haueis, M. Trepte, C. Brenk, A. Tamke, M. Ghanaat, M. Braun, A. Joos, H. Fritz, H. Mock, M. Hein, and E. Zeeb. Making bertha drive - an autonomous journey on a historic route. *IEEE Intelligent Transportation Systems Magazine*, January 2014. Cited on page 1.



Upphovsrätt

Detta dokument hålls tillgängligt på Internet — eller dess framtida ersättare — under 25 år från publiceringsdatum under förutsättning att inga extraordinära omständigheter uppstår.

Tillgång till dokumentet innebär tillstånd för var och en att läsa, ladda ner, skriva ut enstaka kopior för enskilt bruk och att använda det oförändrat för icke-kommersiell forskning och för undervisning. Överföring av upphovsrätten vid en senare tidpunkt kan inte upphäva detta tillstånd. All annan användning av dokumentet kräver upphovsmannens medgivande. För att garantera äktheten, säkerheten och tillgängligheten finns det lösningar av teknisk och administrativ art.

Upphovsmannens ideella rätt innefattar rätt att bli nämnd som upphovsman i den omfattning som god sed kräver vid användning av dokumentet på ovan beskrivna sätt samt skydd mot att dokumentet ändras eller presenteras i sådan form eller i sådant sammanhang som är kränkande för upphovsmannens litterära eller konstnärliga anseende eller egenart.

För ytterligare information om Linköping University Electronic Press se förlagets hemsida <http://www.ep.liu.se/>

Copyright

The publishers will keep this document online on the Internet — or its possible replacement — for a period of 25 years from the date of publication barring exceptional circumstances.

The online availability of the document implies a permanent permission for anyone to read, to download, to print out single copies for his/her own use and to use it unchanged for any non-commercial research and educational purpose. Subsequent transfers of copyright cannot revoke this permission. All other uses of the document are conditional on the consent of the copyright owner. The publisher has taken technical and administrative measures to assure authenticity, security and accessibility.

According to intellectual property law the author has the right to be mentioned when his/her work is accessed as described above and to be protected against infringement.

For additional information about the Linköping University Electronic Press and its procedures for publication and for assurance of document integrity, please refer to its www home page: <http://www.ep.liu.se/>