Aalto University
School of Science
Master's Programme in Computer, Communication and Information Sciences

Jacopo Losi

Structured light assisted real time stereo photogrammetry for robotics and automation

Novel implementation of stereo matching

Master's Thesis
Espoo, 2em

DRAFT! — April 20, 2020 — DRAFT!

Supervisors:	Professor Juho Kannala, Aalto University
	Professor Nicola Conci, University of Trento
Advisor:	Sami Ruuskanen M.Sc. (Tech.)

Aalto University
School of Science
Master's Programme in Computer, Communication and Information Sciences

ABSTRACT OF MASTER'S THESIS

Author:	Jacopo Losi		
Title: Structured light assisted real time stereo photogrammetry for robotics and automation Novel implementation of stereo matching			
Date:		**Pages:**	v + 19
Major:	Autonomous Systems	**Code:**	
Supervisors:	Professor Juho Kannala Professor Nicola Conci		
Advisor:	Sami Ruuskanen M.Sc. (Tech.)		

The abstract provides goal, motivation, background, and conclusions of the work. It has to fit to one page together with the bibliographical information.

If the thesis is in English and the language of school education is Finnish or Swedish, the abstract is written in English and in Finnish or in Swedish. If the language of school education is other than Finnish or Swedish, the abstract is written in English only.

The thesis example file (`thesis-example.tex`), all the chapter content files (`1introduction.tex` and so on), and the Aalto style file (`aalto-thesis.sty`) are commented with explanations on how the Aalto thesis works. The files also contain some examples on how to customize various details of the thesis layout, and of course the example text works as an example in itself. Please read the comments and the example text; that should get you well on your way!

In the thesis template, you can find the text of the abstract in the abstract in the `thesis-example.tex` file, together with the bibliographical information of the abstract tables. !FIXME **This is an example how to use fixme: add your abstract here.** FIXME! Fixme is a command that helps you identify parts of your thesis that still require some work. When compiled in the custom `mydraft` mode, text parts tagged with fixmes are shown in bold and with fixme tags around them. When compiled in normal mode, the fixme-tagged text is shown normally (without special formatting). The draft mode also causes the "Draft" text to appear on the front page, alongside with the document compilation date. The custom `mydraft` mode is selected by the `mydraft` option given for the package `aalto-thesis`, near the top of the `thesis-example.tex` file.

The instructions on how to compile LaTeX *.tex files to *.pdf files like this are giving in the `thesis-example.tex` file as comments and also in this pdf in a Section **??**.

Keywords:	stereo vision; matching cost; census transform; hamming distance; binary pattern; semi-global matching
Language:	English

Acknowledgements

I wish to thank all students who use LaTeX for formatting their theses, because theses formatted with LaTeX are just so nice.

Thank you, and keep up the good work!

Espoo, 5mm Jacopo Losi

Abbreviations and Acronyms

2k/4k/8k mode	COFDM operation modes
3GPP	3rd Generation Partnership Project
ESP	Encapsulating Security Payload; An IPsec security protocol
FLUTE	The File Delivery over Unidirectional Transport protocol
e.g.	for example (do not list here this kind of common acronymbs or abbreviations, but only those that are essential for understanding the content of your thesis.
note	Note also, that this list is not compulsory, and should be omitted if you have only few abbreviations

Contents

Chapter 1

Introduction

1.1 Problem statement

Dense and accurate disparity maps are the key factor for obtaining correct depth estimations for many computer vision applications such as autonomous driving, 3D reconstruction, object detection and robotics. Therefore, stereo matching and disparity estimation can be identified as fundamental problems in the current developments of computer vision [1].

Multiple methods for disparity estimation has been developed for many years [1]. Specifically, older strategies are focused on local-based or global-based methods. On the contrary, deep learning based strategies applied to local or global methods has been recently proposed. The latter approach aims to a precise local correspondence exploiting deep learning and applying Semi-global matching (SGM) as the regularization step of the pipeline. Therefore, deep learning techniques such as FlowNet and DispNet [1] are used as the end-to-end part of the pipeline. According to the current benchmark database ranks for stereo matching algorithms, e.g. the one published in the KITTI website, the state of the art implementations are based on deep learning methods. However, these strategies lack in accuracy compared to the standard pipelines. This is probably due to the difference between real environment and the training database as underlined in [1] [2].

As aforementioned, the state of the art methods to recover dense disparity maps from stereo pairs are focused on deep convolutional neural networks trained end-to-end [3]. Most of these techniques, which will be subsequently described, exploit as regularization phase the Semi-global matching (SGM) method. Actually, among local and global approaches, the Hirschmuller's algorithm [4] appears to be the best performing in terms of computational cost and accuracy. For this reason, it is the preferred trade-off for most real

time applications.
Considering the multiple algorithm for stereo correspondence, they can be conventionally classified [5] into two general categories, local and global approaches. Specifically, the local-based methods tend to estimate the disparity image trough a comparison of the matching cost from left and right views of the scene. In order to recover from low accuracy proper of the previous strategy, global-based methods try to calculate the disparity values by minimizing an energy function. In this context, Semi-Global Matching (SGM) combines strong factors of global and local approaches allowing to obtain a good trade-off between computational cost and accuracy.
Technically speaking, SGM applies a pixelwise, Mutual Information (MI) based matching cost for analysing pixel intensity value differences of input images [4]. Moreover, pixelwise matching is enhanced with a smoothness constraint, which leads to a global cost function. Then, post-processing techniques are applied to remove outliers and filter the image.
Referring to the analysis performed by Scharstein and Szeliski [5], SGM carries out four main steps, as well as most of the stereo matching algorithms. These are defined as: matching cost computation, cost aggregation, disparity computation and disparity refinement.
Considering the former, it is usually based on absolute, squared or sampling insensitive difference between pixel intensities [4]. Although those methods allow to reach a reliable accuracy, they are sensitive to radiometric difference. Thus, cost based on image gradients or window-based methods, such as rank and census transform [6], became an optimal choice. Furthermore, Mutual Information results as a good trade-off for dealing with complex radiometric relationships between images.
In the second phase, cost aggregation collects the matching costs considering multiple directions and the disparity levels. Following, disparity evaluation is defined for each pixel, as the one with the lowest cost. This is the approach typically used for local methods. Global algorithms, rather, used to get rid of the aggregation step and define a global energy function. Over that function, pixel similarity and disparity smoothness are enforced with different strategies. In this latter case, the best disparity is identified finding the minimum of the cost function. This is achieved with multiple techniques such as: Dynamic Programming (DP) [7], Belief Propagation [8] or Graph Cuts [9].
Disparity refinement tends to differ more among the different methods. Usually, post-processing techniques such as filtering, outlier removal and consistency check are in general applied.
As anticipated above, among the top-ranked stereo matching algorithms, SGM results to be the best performing in terms of computational time and

accuracy. Its benefits stand in the hierarchical computation of the matching cost, which exploit Mutual Information. Cost aggregation is achieved taking into account a global energy function and a pathwise pixel optimization. The final disparity is chosen with a winner takes all strategy. Disparity refinement is completed by consistency check between left and right disparity images.
Besides the challenge of building up the optimal algorithm for recovering a disparity image from a stereo image pair, it is necessary to develop an analysis of the basis of stereo correspondence and its importance for multiple applications such as: autonomous driving, robotics, object detection and 3D reconstruction.
First of all, stereo matching is defined as the process of estimating a 3D model of a scene, starting from two or more images. Therefore, the matching pixel between the images are found and their 2D positions are converted into 3D depths. Thus, how this operation of building a dense depth map, assigning relatives depth to the input image pixels, is achieved. This is based on the disparity, defined as the amount of horizontal motion between two properly configured images of a stereo pair. This one is then inversely proportional to the distance from the observer, i.e. the camera. Although this concepts are relatively simple to understand, the challenging task within this process stands in establishing dense and accurate inter-image correspondences[10].
As already underlined, stereo matching is one of the most widely studied topic in computer vision from years and it continues to be one of the most active research in that field. In fact, modelling of human visual systems, robotic navigation and manipulation and autonomous driving [2] and 3D model building are some of the possible applications.
The explanations of the fundamental principles of stereo matching, such as epipolar geometry, rectification and disparity map, follows.

1.1.1 Stereo geometry

Main goal of epipolar geometry is the computation of pixels correspondences among the input images. Neighbouring pixels information, cameras positions and their calibration data are fundamental to achieve that. Figure 1.1 demonstrate a pixel in one image $\mathbf{p}_1$ projected to its correspondent epipolar line segment in the other image, which is lower bounded by the projection of the first camera center into the second camera plane, i.e. the epipole $\mathbf{e}_2$. Projecting the epipolar line in the second image back to the first, another line would be obtained, bounded by the correspondent epipole $\mathbf{e}_1$. The extensions to infinity of these two segment are identified as the epipolar lines, which are defined by the intersection of the two image planes with the epipo-

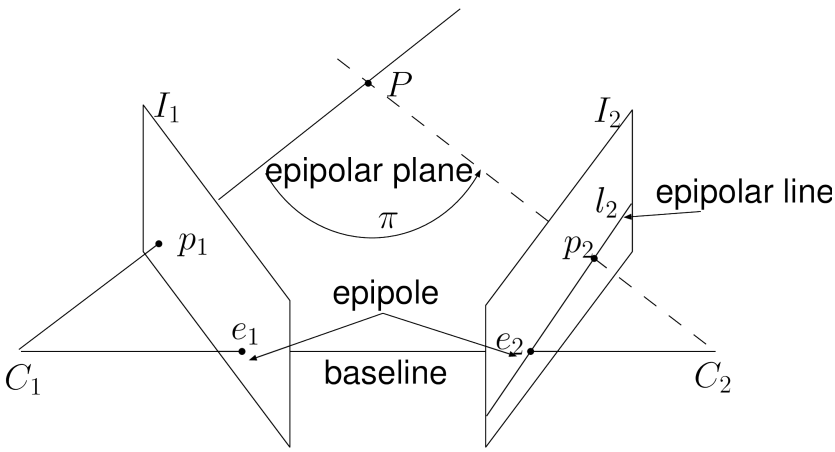

Figure 1.1: Epipolar geometry

lar plane. A fundamental property is that the epipolar plane passes through both camera centers $\mathbf{C}_1$ and $\mathbf{C}_2$, as well as point $\mathbf{P}$. Therefore, they lie in the same plane.

1.1.2 Rectification

Epipolar geometry for a pair of cameras is relative to pose and calibration of the camera and can be computed using the fundamental matrix, which can be obtained applying the eight point algorithm [11]. Computing this geometry allows, then, to find the correspondent pixels between the two images using the constraint of the epipolar lines. This is possible, because, as explained in 1.1.1, considered a pixel in one image, the correspondent one lies on the relative epipolar line.
Beside that, pixels correlations can be more efficiently performed by rectifying the input images [11]. In Figure 1.2 is clearly visible the outcome of this process and its advantages. As shown, corresponding horizontal scanlines are epipolar lines. The essential importance of this standard rectified geometry is clearly explained by the following equation,

$$d = f\frac{B}{Z} \tag{1.1}$$

that leads to a linear relationship between 3D depth Z and disparity d, where f is the focal length (in pixel) and B the baseline. Moreover, the relationship between the corresponding pixels in the left and right images

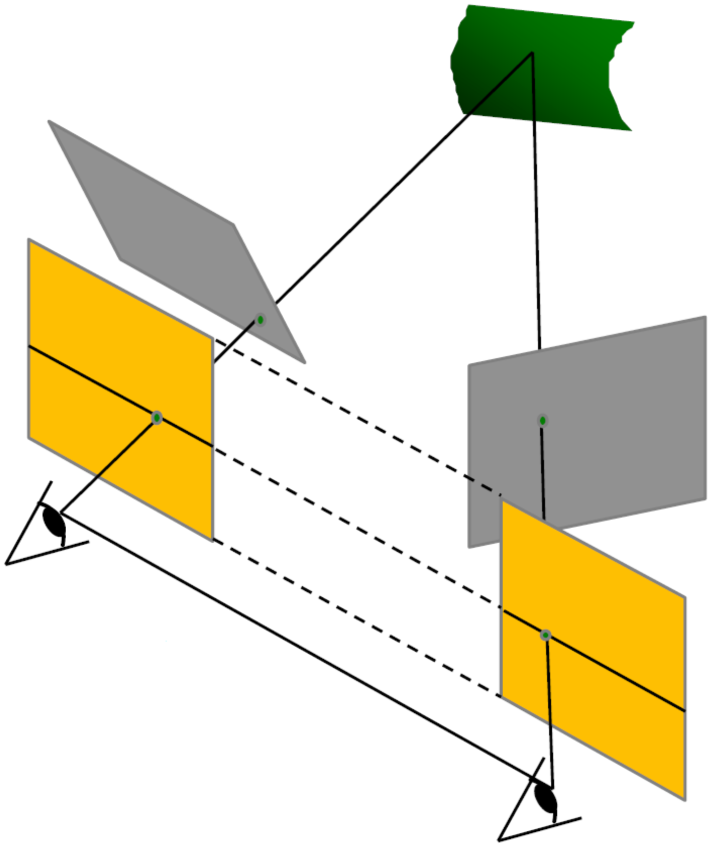

Figure 1.2: Image rectification – Source: L. Lebeznik

can be defined as follows:

$$x' = x + d(x, y),\ y' = y \tag{1.2}$$

Thus, the main step for recovering a depth image of a scene is the estimation of the disparity map $d(x, y)$.
As introduced at the beginning, the best disparity map is estimated after the rectification process. This is performed by comparing the similarity of corresponding pixels, as defined in equation 1.2, and storing them in a disparity space image $C(x, y, d)$, which is then processed with multiple algorithms.

1.1.3 Stereo methods and dense correspondence

In this section a brief delineation of the general pipeline implemented in most of the stereo matching method is presented. Moreover, as a theoreti-

cal completion of what introduced above, some generic algorithm are further explained.
Stereo algorithm follows in general a subspace of the following methods: matching cost computation, cost aggregation, disparity computation and optimization and disparity refinement [5].
A preliminary distinction, based on those phases, separates stereo methods between local or window-base global methods.
In local methods, the disparity computation in a certain region depends on the pixel intensities within a limited window.
On the contrary, global algorithms, are based on an energy function. In this one smoothness assumptions are defined and then a global optimization problem is solved. These algorithm are mainly distinguished considering the minimization strategy, such that, simulated annealing, graph cuts or belief propagation.
Between these two classes there are iterative and hierarchical algorithms. The latter aim to constraint gradually the disparity estimation from the coarser to the finer levels [4].
Considering the first general step of stereo matching algorithms, the matching cost, there are multiple measures to define it. Among the most prevalent pixel-based algorithm can be included square intensity differences, absolute intensity differences, mean-squared error and mean absolute difference.
Other common matching cost comprehend normalized cross-correlation, which is similar to sum-of-squared-difference (SSD), and binary methods. However, the latter tend to not be used any longer.
On the other hand, lately, more robust algorithms are used for their insensitivity to non-stationary exposure and illumination changes. Entropy measures and non-parametric functions such as, rank and census transform [12], sampling insensitive difference[7] and hierarchical mutual information [4], are some examples. In particular, they allow to obtain accurate performance when considerable exposure or appearance variations are present.
Drawing up some conclusion regarding the local methods, the core steps are the matching cost calculation and the aggregation phase. Disparity estimation, then, becomes trivial. Each pixel takes the disparity levels whose cost value is the minimum. This approach is said to be a local *winner-take-all* optimization. A drawback of this approach is that the matches are imposed for the reference image. While points is the support image might have multiple correct matches. For this reason, cross-checking and post-processing become more relevant here.
Summarizing the general pipeline of global stereo matching methods, they often get rid of the aggregation step. They usually perform sort of optimization steps after disparity estimation, exploiting the smoothness constraints

as aggregation part.
Goal of this approach is to find the solution to a global energy function, i.e. the disparity d that minimizes the energy,

$$E(d) = E_d(d) + \lambda E_s(d) \tag{1.3}$$

where $E_d(d)$ is the data term and $E_s(d)$ the smoothness term. Adopting the aforementioned disparity space image (DSI) matching cost, the data energy is calculated as:

$$E_d(d) = \sum_{(x,y)} C(x, y, d(x, y)) \tag{1.4}$$

where C is the DSI. Then, the smoothness term is usually defined as:

$$E_s(d) = \sum_{(x,y)} \rho(d(x, y) - d(x + 1, y)) + \rho(d(x, y) - d(x, y + 1)) \tag{1.5}$$

where ρ is some monotonically increasing function of disparity difference. For some implementations, the smoothness energy term can also be based on intensity differences,

$$\rho_d(d(x, y) - d(x + 1, y)) \cdot \rho_I(\|I(x, y) - I(x + 1, y)\|) \tag{1.6}$$

where ρ_I is a monotonically decreasing function, which depends on the intensity differences and makes the smoothness costs lower at high-intensity gradients.
After the energy function has been clearly identified, different categories of algorithms can be exploited to recover a (local) minimum. Graph cut, belief propagation and Markov random field (MRF) based methods have been proved to give the most accurate results.
Mentioning some hybrid methods, there are cooperative algorithms and others based on coarse to fine incremental steps. Cooperative algorithms were some of the earliest proposed for disparity estimation. They are influenced by human stereo vision processing models. They operate by iteratively improve disparity evaluations using non-linear calculations, leading to a result similar to the one of the global methods. Iterative algorithms are, instead among the current best algorithms. The main idea is to successively choose the best disparity among all of the possible ones. A coarse-to-fine framework is usually used to speed up the computations.
Dealing with global optimization methods, it is worth to mention dynamic programming (DP) technique. Unlike solutions based on equation 1.3, dynamic programming allows to reach global minimum exploiting independent scanlines. This approach works over a slice of the DSI, i.e. the matching

cost cube, finding the path associated to the minimum cost. The generic implementation of DP along a scanline y and for each input state in a 2D cost matrix $D(m,n)$ leads to combine its DSI value with its previous cost values as follows,

$$C'(m,n) = C(m+n, m-n, y) \tag{1.7}$$

Correct cost selection in presence of occluded pixels and difficulties with scanline consistency are some of the weakness of DP. Multiple algorithms have been proposed to recover from these problems. Scharstein and Szeliski [5] proposed an alternative to standard DP, improving recursively independent scanlines in the global energy function,

$$D(x,y,d) = C(x,y,d) + \min_{d'}\{D(x-1,y,d') + \rho_d(d-d')\} \tag{1.8}$$

An upgrade of this scanline optimization is the aggregation cost approach used in semi-global matching [4]. In this case, a cumulative cost function is evaluated from at least eight directions. Intuitively, this approach accesses accurate results and it is highly efficient.
Considering more recent improvements to stereo matching, segmentation-based techniques hold a prominent position. In this case, an initial segmentation of the reference image is performed. Then, disparities are estimated pixelwise using local methods. Citing couple of recent approaches, Klaus, Sormann and Karner [8] segment the image with mean shift, to get initial disparity estimations. Then they apply re-fitting with global planes, and perform final MRF with loopy belief propagation.
Wang and Zheng [**?**] built a similar top ranked algorithm. They segment the image with local plane fits. Then run cooperative optimization of neighbouring plane fit parameters. Others developed similar approaches exploiting color correlation and left-right consistency check for occlusion detection [13] or focusing on alpha mattes fractional pixel extraction [14].
As explained in section 1.1 the area of stereo matching and disparity estimation is one of the most extensively researched topic in computer vision. Nowadays approaches based on convolutional neural networks and deep learning are going to be the highly ranked in the standard database. Although, their performance in accuracy tends to decrease a lot when moving from dataset images to real ones.
Therefore, novel strategies based on standard stereo geometry algorithms could reach consistent accuracy even in real time [15]. Thus, as described above, after the structure of the cost volume or DSI has been delineated, the actual pixelwise photoconsistency measures are computed. Multiple methods to achieve this has been proposed during years and already explained in the previous paragraphs. Then, depth computation is obtained with different

form of optimizations. These ranges among local, global or hybrid frameworks.
Consequently, starting from classical stereo-based methods and building up a novel pipeline, accurate real time depth map estimations can be achieved.

1.2 Structure of the Thesis

You should use transition in your text, meaning that you should help the reader follow the thesis outline. Here, you tell what will be in each chapter of your thesis. Often the thesis does not have as many chapters as is in this template. For example, environment and implementation can be combined as well as chapters of evaluation and discussion. The rest of this thesis is organized as follows. Chapter 2 gives the background, etc.

Chapter 2

Theoretical Background and Related Work

In chapter 1 a brief general analysis of stereo geometry and methods has been provided. In this chapter a more precise revision of the theoretical tools that stereo matching methods exploit is presented. Epipolar geometry, camera calibration and disparity estimation algorithms are specifically described. Starting from the necessary mathematical basis, the discussion moves on the disparity estimation algorithms. Then, the chapter focus on the main benefits and drawbacks of standard and novel approaches in depths computation. Comparison between stereo-geometry based and deep learning based algorithms is proposed, to provide a clear explanation of the decisions implemented.

2.1 Epipolar geometry and Rectifiation

Fundamental problem of stereo vision is the estimation of 3D locations of points from at least two corresponding input images. This process, which comprises concurrent computation of both 3D geometry and camera pose, is generally known as structure from motion [10].
In the explanation of these topics it is necessary to start discussing about the triangulation. Then, the concept of epipolar geometry is outlined and after that the notions of camera calibration and rectification.

2.1.1 Triangulation

Triangulation is the problem of detecting 3D points positions from a collection of corresponding 2D image locations, assuming that camera positions

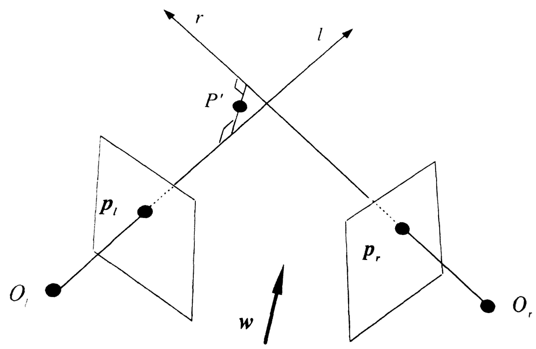

Figure 2.1: 3D triangulation by finding point P' that lies nearest to all of the optical rays

are known. Figure 2.1 shows one of the easiest methods to tackle this problem. Objective is to evaluate the 3D position of P' that have the smallest error to all of the 3D optical rays coming from the camera centers, which identify the 2D point locations in the image plane, i.e. P_r and P_l. As shown in Figure 2.1, the rays starts from the camera centers, O_j and go in direction of r and l, which can be defined using the camera matrix $\{P_j = K_j[R_j|t_j]\}$. The closest point to P on this ray minimizes the distance

$$\|O_j + d_j\hat{v}_j - P\|^2 \tag{2.1}$$

Therefore, because of the minimum is $d_j = \hat{v}_j \cdot (p - c_j)$, the nearest points are calculated as:

$$q_j = O_j + (\hat{v}_j\hat{v}_j^\top)(P - O_j) = O_j + (P - O_j)_\| \tag{2.2}$$

Hence, the optimal value for P, obtained solving a least square problem, becomes,

$$P = \Big[\sum_j (I - \hat{v}_j\hat{v}_j^\top)\Big]^{-1} = \Big[\sum_j (I - \hat{v}_j\hat{v}_j^\top)O_j\Big] \tag{2.3}$$

2.1.2 Epipolar geometry

The intrinsic projective geometry between two views is known as epipolar geometry. It is only dependent on the cameras' internal parameters and pose. The 3×3 rank 2 matrix that defines this geometry is the fundamental matrix F.
The epipolar geometry is the basis for finding corresponding points in stereo

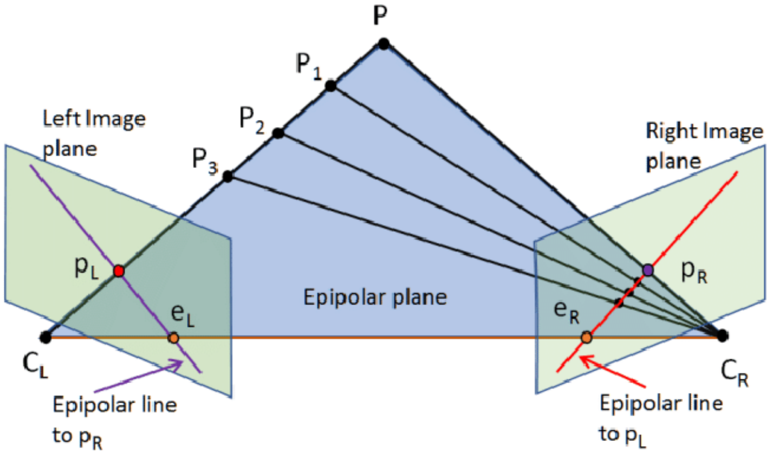

Figure 2.2: Epipolar geometry. Image point m back-projects to a ray in a 3D space defined by C and m. This ray becomes a line l' in the second view. The image of X must lie on l'

matching. It is basically defined by the intersection between image planes and the one on which the cameras baseline lies.
A fundamental property, that makes this geometry extremely useful, is that image points, space point and camera centers are coplanar. Assuming that only m_l is known, that geometry allows to constraint the corresponding point m_r. The epipolar plane is defined by the baseline and the ray that comes from m_l. Hence, knowing that m_r lies on the same plane, that point belongs to l_r, i.e. the intersection between the epipolar and the second image plane. Therefore, exploiting this property, the searching of corresponding points is constrained to only one line inside the image.
Mathematical definition of the epipolar geometry is the fundamental matrix F. As already demonstrated through Figure 2.2, for each point p_L in one image, the corresponding epipolar line e_R to that point belongs to the other image plane. Moreover, any point p_R in the second image, which is related to point p_L, lies on e_R. Hence, the epipolar line is described as the projection in the second image of the ray that comes from the point in the first image, passing through its camera center. This defines a map, $p_L \rightarrow e_R$, which relates the points in one image with the corresponding epipolar lines in the second image. This correlation, between points and lines, is represented by the fundamental matrix F.
Considering the aforementioned map $p_L \rightarrow e_R$ described by F, an important

property of the fundamental matrix is defined,

$$p_R^\top F p_L = 0 \tag{2.4}$$

Therefore, assuming two corresponding points P_L and p_R, it is known that p_R lies on the epipolar line $l_R = Fp_L$. Thus, the mathematical correlation is,

$$0 = p_R^\top e_R = p_R^\top F p_L \tag{2.5}$$

Reciprocally, if image points comply the relation 2.4, then the rays identified by these points are coplanar. For point corresponding this is a necessary condition. Equation 2.4 is extremely important because it allows to characterize the fundamental matrix without reference to the camera matrices [11]. Thus, using at least 7 correspondences, it is possible to recover the fundamental matrix F. This estimation is known as *weak calibration*.
Maybe add 8-point algorithm description

2.1.3 Rectification

Image rectification is defined as the process of obtaining a pair of *matched epipolar projections* from a pair of stereo images, which are taken from generally differing viewpoints. In the rectified projections the epipolar lines become parallel with respect to the x-axis. Thus, they match between the stereo pair and so the disparities are in the x-direction only.
In order to obtain a rectified stereo pair, 2D projective transformations are employed to the images, so that the epipolar lines can match. Using this method, the transformations are built up in a way that the corresponding points have almost the same x-coordinate. Actually, this strategy leads to a minimal distortion on the images, being the two transformations arbitrary. However, working on rectified images, the matching problem is highly simplified, being correlated only to epipolar geometry and near-correspondence. Core problem of this section is to find the appropriate projective transformation H. Indeed, to get epipolar lines parallel with x-axis, the epipole should be mapped to an infinite point. This, has to be done correctly, otherwise intensive projective distortion of the image can happen. For this reason, constraints are put on the definition of H.
First of all, restricting H to be a rigid transformation in the neighbourhood of a given point[1], the errors are reduced.
Once the epipole has been mapped to infinity, it is then necessary define a

[1]this means that to first-order, the neighbourhood of the point may be subjected to rotation and translation only

map to match the corresponding epipolar lines. This resampling is build up in such a way that, being e_L and e_R any pair of epipolar lines, then,

$$H^{-\top} e_L = H'^{-\top} e_R \tag{2.6}$$

Satisfying the condition above, a matched pair of transformations is recovered.
Specifically, at first H' is chosen, so that it can map the epipole e_R to infinity. Then the matching transformation H is defined minimizing the sum-of-square distances,

$$\sum_i d(Hp_{L_i}, H'p_{R_i})^2 \tag{2.7}$$

Therefore, the full algorithm can be summarized as follows.
The outcome of this resampling process is a pair of stereo images whose epipolar lines are horizontal. Hence, the disparities are calculated along the epipolar lines. First of all, at least seven corresponding matches are defined. This allows to compute the fundamental matrix F, applying the so called eight-point algorithm, and after that the two epipoles are found. After that, there is the selection of the projective transformation H', that maps the epipole of the support image to infinity. The corresponding transformation H is found solving the least-square problem. Finally both of the input images are resampled according to H and H'.

2.2 Stereo methods and dense correspondence

Take the part already written in the introduction
Refactoring needed between the 2 chapters

2.2.1 Stereo geometry based methods

2.2.2 Deep learning based methods

2.2.3 Referring to sources

*Haapasalo [**?**] researched database algorithms that allows use of previous versions of the content stored in the database.* This kind of marking means that this paragraph (or until the next reference is given) is based on the source mentioned in the beginning. Giving the source you should use only the family name of the first author of the article, and not give any hints about what is the type of the article that is referred nor its title.

B+-trees offers one way to index data that is stored in to a database. Multiversion B+-trees (MVBT) offer also a way to restore the data from *previous versions of the database. Concurrent MVBT allows many simultaneous updates to the database that is was not possible with MVBT. [**?**]* When the marking is after the period, the reference is retrospective: all the paragraph (or after previous reference marking) is based on the source given in its end. If the content is very broad, you can start with saying *According to Haapasalo*, then continue referring the source with several separate sentences, and in the end put the marking of your source *that shows that CMVBT are the best. [**?**].*

If your paragraph has several sources, the above mentioned styles are not proper. The reader of your thesis cannot know which of your sources give which of the statements. In this case, it is better to use more finegraded refering where the reference markings that are embedded in the sentences. For example, *the multiversion B+-tree (MVBT) index of Becker et al. [**?**] allows database users to query old versions of the database, but the index is not transactional. It's successor, the transactional MBVT (TMVBT), allows a single transaction running in its own thread or process to update the database concurrently with other transactions that only read the database [**?**]. Further development, titled the concurrent MBVT (CMVBT), allows several transactions to perform updates to the database at the same time [**?**].* Here, the references are marked before the period in the sentences where they are used. You should never but all these sources in the end of the paragraph. Referring several source at once should only used when you give a set of examples.

Finally, direct quotes are allowed. However, often you should avoid them since they do not usually fit in to your text very well. Using direct quotes has two tricks: quotation marks and the source. *"Even though deletions in a multiversion index must not physically delete the history of the data items, queries and range scans can become more efficient, if the leaf pages of the index structure are merged to retain optimality." [**?**]* Quotes are hard to make neatly since you should use only as much as needed without changing the text. Moreover, you often do not really understand what the author has mentioned with his wordings if you cannot write the same with your own words. Remember also that never cut and paste anything without marking the quotation marks right away, and in general, never cut and paste anything at all!

Sometimes getting the original source can be almost impossible. In an extremely desperate situation, you can refer with structure *ms X [...] according to mr Y [...] defined that*, if you find a source that refers to the original source. Note also that the reference marking is never used as sentence element (example of how **not** to do it: *[**?**] describes an optimal algorithm for indexing multiversiond databases.*).

Bibliography

[1] A. Seki, M. Pollefeys, T. Corporation, E. T. Zürich, and Microsoft, "SGM-Nets: Semi-global matching with neural networks," *Proceedings - 30th IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017*, vol. 2017-Janua, no. 1, pp. 6640–6649, 2017.

[2] M. Poggi, D. Pallotti, F. Tosi, and S. Mattoccia, "Guided Stereo Matching," 2019.

[3] A. Tonioni, F. Tosi, M. Poggi, S. Mattoccia, and L. D. Stefano, "Real-Time Self-Adaptive Deep Stereo," pp. 195–204, 2020.

[4] H. Hirschmüller, "Stereo processing by semiglobal matching and mutual information," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 30, no. 2, pp. 328–341, 2008.

[5] D. Scharstein, R. Szeliski, and R. Zabih, "A taxonomy and evaluation of dense two-frame stereo correspondence algorithms," *Proceedings - IEEE Workshop on Stereo and Multi-Baseline Vision, SMBV 2001*, no. December, pp. 131–140, 2001.

[6] J. Ko and Y. S. Ho, "Stereo matching using census transform of adaptive window sizes with gradient images," *2016 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference, APSIPA 2016*, pp. 2–5, 2017.

[7] S. Birchfield and C. Tomasi, "Depth discontinuities by pixel-to-pixel stereo," *International Journal of Computer Vision*, vol. 35, no. 3, pp. 269–293, 1999.

[8] A. Klaus, M. Sormann, and K. Karner, "Segment-based stereo matching using belief propagation and a self-adapting dissimilarity measure," *Proceedings - International Conference on Pattern Recognition*, vol. 3, pp. 15–18, 2006.

[9] V. Kolmogorov and R. Zabih, "Computing visual correspondence with occlusions using graph cuts," *Proceedings of the IEEE International Conference on Computer Vision*, vol. 2, pp. 508–515, 2001.

[10] R. Szeliski, "Computer vision: algorithms and applications," *Choice Reviews Online*, vol. 48, no. 09, pp. 48–5140–48–5140, 2011.

[11] R. Hartley and A. Zisserman, "Multiple view geometry in computer vision. cambridge university press, isbn: 0521540518," 2004.

[12] R. Zabih and J. Woodfill, "Non-parametric local transforms for computing visual correspondence," *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 801 LNCS, pp. 151–158, 1994.

[13] Q. Yang, L. Wang, R. Yang, H. Stewénius, and D. Nistér, "Stereo matching with color-weighted correlation, hierarchical belief propagation, and occlusion handling," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 31, no. 3, pp. 492–504, 2008.

[14] M. Bleyer, M. Gelautz, C. Rother, and C. Rhemann, "A stereo approach that handles the matting problem via image warping," in *2009 IEEE Conference on Computer Vision and Pattern Recognition*, pp. 501–508, IEEE, 2009.

[15] D. Hernandez-Juarez, A. Chacon, A. Espinosa, D. Vazquez, J. C. Moure, and A. M. Lopez, "Embedded real-time stereo estimation via Semi-Global Matching on the GPU," *Procedia Computer Science*, vol. 80, pp. 143–153, 2016.

Appendix A

First appendix

This is the first appendix. You could put some test images or verbose data in an appendix, if there is too much data to fit in the actual text nicely.

For now, the Aalto logo variants are shown in Figure A.1.

(a) In English

Figure A.1: Aalto logo variants