Queen Mary University of London

School of Electronic Engineering and

Computer Science

Multimedia and Vision Research Group

Doctor of Philosophy Thesis

ACCURATE AND FAST STEREO VISION

by

Georgios Kordelas

Supervisor: Prof. Ebroul Izquierdo

August 2015

Organization:
School of Electronic Engineering and Computer Science, Queen Mary University of London, London, United Kingdom

Title:
Accurate and Fast Stereo Vision

Author:
Georgios Kordelas

Supervisors:
Prof. Ebroul Izquierdo (Queen Mary University of London)
Dr. Ioannis Patras (Queen Mary University of London)
Dr. Pengwei Hao (Queen Mary University of London)

To my parents, Athanasios and Vassiliki,
and my sisters, Elpiniki and Drosoula…

Statement of Originality

I, Georgios Kordelas, confirm that the research included within this thesis is my own work or that where it has been carried out in collaboration with, or supported by others, that this is duly acknowledged below and my contribution indicated. Previously published material is also acknowledged below.

I attest that I have exercised reasonable care to ensure that the work is original, and does not to the best of my knowledge break any UK law, infringe any third party's copyright or other Intellectual Property Right, or contain any confidential material. I accept that the College has the right to use plagiarism detection software to check the electronic version of the thesis. I confirm that this thesis has not been previously submitted for the award of a degree by this or any other university. The copyright of this thesis rests with the author and no quotation from it or information derived from it may be published without the prior written consent of the author.

Signature: Georgios Kordelas
Date: 23/08/2015

Details of collaboration and publications:

All papers published while working on this thesis are listed at the end of the thesis. Any publication produced in collaboration with others is clearly mentioned.

Abstract

Stereo vision from short-baseline image pairs is one of the most active research fields in computer vision. The estimation of dense disparity maps from stereo image pairs is still a challenging task and there is further space for improving accuracy, minimizing the computational cost and handling more efficiently outliers, low-textured areas, repeated textures, disparity discontinuities and light variations.

This PhD thesis presents two novel methodologies relating to stereo vision from short-baseline image pairs:

I. The first methodology combines three different cost metrics, defined using colour, the CENSUS transform and SIFT (Scale Invariant Feature Transform) coefficients. The selected cost metrics are aggregated based on an adaptive weights approach, in order to calculate their corresponding cost volumes. The resulting cost volumes are merged into a combined one, following a novel two-phase strategy, which is further refined by exploiting semi-global optimization. A mean-shift segmentation-driven approach is exploited to deal with outliers in the disparity maps. Additionally, low-textured areas are handled using disparity histogram analysis, which allows for reliable disparity plane fitting on these areas.

II. The second methodology relies on content-based guided image filtering and weighted semi-global optimization. Initially, the approach uses a pixel-based cost term that combines gradient, Gabor-Feature and colour information. The pixel-based matching costs are filtered by applying guided image filtering, which relies on support windows of two different sizes. In this way, two filtered costs are estimated for each pixel. Among the two filtered costs, the one that will be finally assigned to each pixel, depends on the local image content around this pixel. The filtered cost volume is further refined by exploiting weighted semi-global optimization, which improves the disparity accuracy. The handling of the occluded areas is enhanced by incorporating a straightforward and time efficient scheme.

The evaluation results show that both methodologies are very accurate, since they handle efficiently low-textured/occluded areas and disparity discontinuities. Additionally, the second approach has very low computational complexity.

Except for the aforementioned two methodologies that use as input short-baseline image pairs, this PhD thesis presents a novel methodology for generating 3D point clouds of good accuracy from wide-baseline stereo pairs.

Table of Contents

List of Figures

List of Tables

List of Abbreviations

Alphabetically

3D	Three Dimensional
AD	Absolute Differences
BT	Birchfield-Tomasi
CIELab	Commission Internationale de l'Eclairage Lab
CPU	Central Processing Unit
EDISON	Edge Detection and Image SegmentatiON system
GB	Gibabyte
GHz	Gigahertz
GPU	Graphics Processing Units
MLS	Moving Least Squares
PCA	Principal Components Analysis
RAM	Random Access Memory
RANSAC	RANdom SAmple Consensus
RGB	Red Green Blue
SfM	Structure-from-Motion
SIFT	Scale Invariant Feature Transform
SURF	Speeded Up Robust Features
TOF	Time-of-Flight
WTA	Winner-Takes-All

Acknowledgements

The first words of this thesis are dedicated to all people that contributed, in a way or another, to the accomplishment of such an important milestone in my life.

To start, I would like to express my heartfelt gratitude to Prof. Ebroul Izquierdo and Dr. Petros Daras for giving me the opportunity to pursue this PhD at the Queen Mary University of London. I am also grateful for their important advices and guidance during my PhD research. Additionally, I owe special thanks to Dr. Dimitrios Alexiadis for his cooperation and his advices. I would also like to thank Dr. Ioannis Patras and Dr. Pengwei Hao for providing me valuable advices through the stages of this PhD. Also, I am indebted to my thesis examiners Prof. Panos Liatsis and Prof. Jonathan Freeman for giving insightful remarks and comments, which were vital for improving the quality of this thesis.

A big thanks goes to all my friends and family that, even without noticing, contributed to this goal. It was, many times, that their words and actions made me face my worries with a much more positive attitude.

Chapter 1

1. Introduction

Computer vision is a scientific field that studies how to reconstruct, interpret and understand a real word scene from its 2D images in order to extract numerical or symbolic information. This information may be exploited in a second stage by an autonomous system to take proper decisions and perform proper actions.

This PhD thesis deals with stereo vision, which is an important sub-domain of computer vision. Stereo vision aims at estimating the 3D structure of a scene from two images that capture the scene from two different viewpoints. The 3D information can be extracted by detecting the relative position of the scene's objects in the two images.

The process of retrieving 3D information from stereo cameras may appear simple, since humans are able to perceive 3D data naturally using human binocular vision. However, it turns out that is a complex task for a computer.

In the following of Chapter 1, after mentioning several applications of stereo vision, the background and the research objectives relevant to stereo vision are provided. Additionally, this Chapter concisely describes the contributions of this PhD towards the realization of the research objectives, before giving the outline of this thesis.

1.1 Applications of stereo vision

Stereo vision, which is one of the most active research fields in computer vision [1], is exploited in a wide range of applications such as:

- **Robot navigation**: Autonomous robot navigation in dynamic environments requires the study of relative motion of the objects in the robot's environment with respect to the robot and the analysis of depth towards those objects. Stereo vision can be used to efficiently estimate the depth to the surfaces that lie in the vicinity of the mobile robots [2], [3].

- **Augmented reality:** Stereo vision processing is a critical component of augmented reality systems that rely on the precise depth map estimation of a scene to properly place computer generated objects with real life video [4], [5].
- **Automotive applications:** The 3D perception of a car's surroundings is crucial, both for driver assistance and for safety systems. An option to obtain 3D measurements of the surroundings is to use a stereo vision system [6], [7].

1.2 Background of stereo vision and disparity estimation

1.2.1 Stereo vision theory

This subsection provides the basic theory of stereo vision [8]. In stereo vision, two cameras, displaced horizontally from one another are used to obtain two differing views on a scene. By comparing these two images, the relative depth information can be obtained. In more detail, the two images are shifted together over top of each other to find the parts that match. The shifted amount is called "disparity''. The higher is the disparity of an object pixel the closer is this object to the cameras. If the object lies very far from cameras, the disparity is approximately zero. This means the object on the left image is the same pixel location as on the right image.

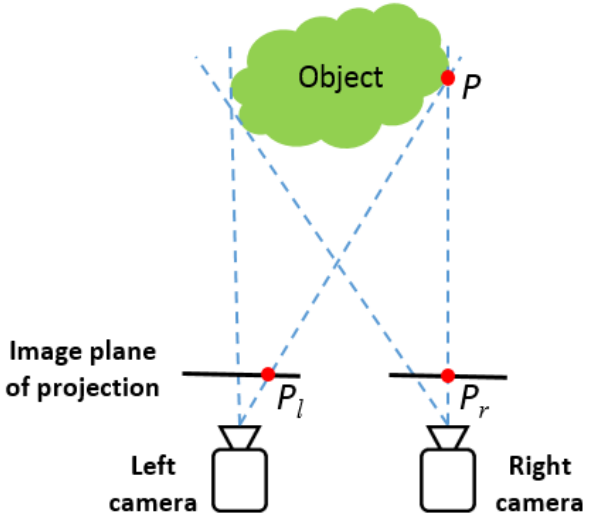


Figure 1. A stereo pair of images and their image planes of projection.

Figure 1 depicts an example of a stereo vision system. The stereo cameras lie on the same plane and have the same direction. The position of both cameras is different

along the X-axis. The image planes are presented in front of the cameras in order to ease the modeling of the projection.

Now, let us consider a point *P* on a scene's object, whose perspective projections are located at pixels P_l and P_r on the image planes of left and right cameras, respectively. The left camera's projection point P_l is shift from the center, while the right camera's projection point P_r is at the center. This shift between the corresponding pixels on the left and the right camera images should be computed to get the depth information of the object.

Therefore, the main purpose of stereoscopic vision is to estimate the corresponding (matching) pixels between the left and right image of the stereo image pair. Afterwards, the depth information can be generated from the corresponding points.

In order to limit the search pixel-correspondences along the same horizontal epipolar line the stereo image pairs have been rectified to epipolar geometry. The epipolar geometry of stereoscopic camera is illustrated in Figure 2. This simple stereo model shows two different perspective views of an object point *P* from the two cameras centers F_l and F_r, which separate only in the x-axis direction by a baseline distance. Points P_l and P_r, which are the perspective projections of P in left and right view, constitute a pair of corresponding pixels. The plane passing through the camera centers F_l, F_r and the object point *P* in the scene is called the epipolar plane. The intersection of the epipolar plane with each image plane is called epipolar line and projections at pixels P_l and P_r lie on the same epipolar line.

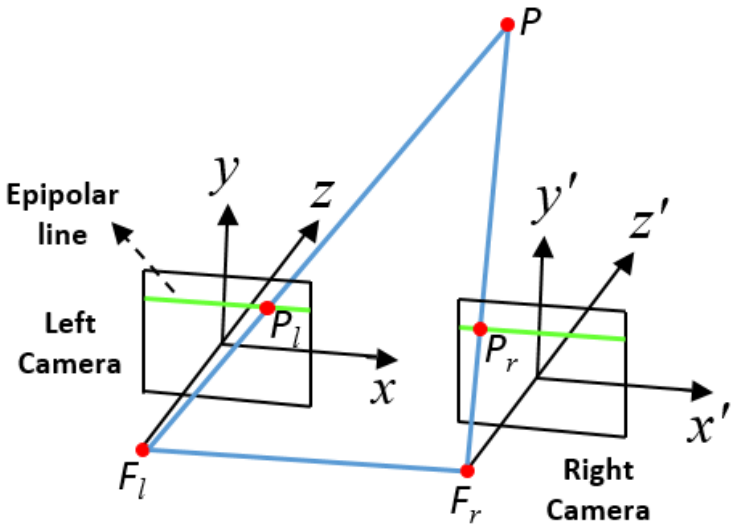


Figure 2. The epipolar geometry of stereo vision.

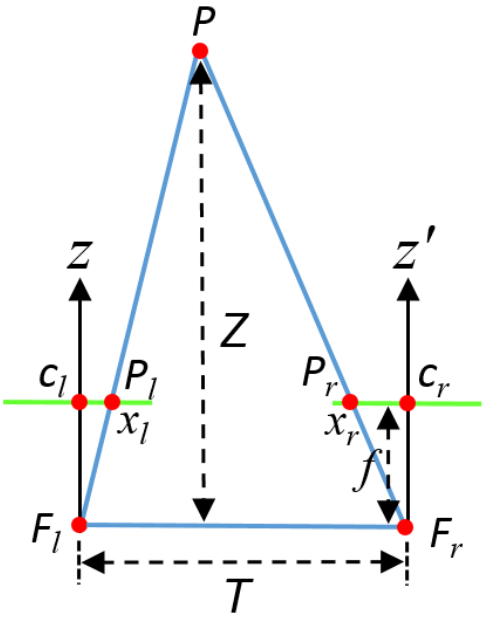


Figure 3. Top view of the epipolar geometry.

The matching of pixels between the two views is the standard procedure for the depth recovery. The depth information can be evaluated by using the triangle similarity algorithms as follows. Figure 3 displays the top view of the epipolar geometry of Figure 2. The distance, T , between the two camera centers F_l and F_r, is called the baseline of the stereo system and f is the focal length. While, x_l and x_r are the coordinates of P_l and P_r, with respect to the principal centers c_l and c_r . The distance Z of point P to the baseline can be determined by comparing the similar triangles (P,F_l ,F_r) and (P,P_l,P_r). After straightforward computations, distance Z is given by:

$$Z = f\frac{T}{x_l - x_r} = f\frac{T}{d}, \qquad (1)$$

where the disparity is: $d = x_l - x_r$.

The main objective of this PhD is to improve the process of estimating of dense disparity maps (i.e. to estimate for all pixels of a stereo image their corresponding pixels on the other image). Dense disparity maps can be used to compute, in a second stage, the depth information via Equation (1).

1.2.2 Disparity estimation steps

The process of estimating disparity maps can be roughly divided into four steps [1]: matching cost computation, cost aggregation, disparity optimization, and disparity refinement.

The matching cost computation step deals with the definition of cost metrics that measure the similarity of two corresponding pixels from the left and right images, respectively. The cost aggregation step relies on supporting pixel areas (i.e. pixel neighbourhoods) to aggregate pixel-based matching costs in order to reduce the ambiguity of matching. The disparity optimization step is used then to correct errors in the disparity maps, which are acquired after performing the matching cost computation and cost aggregation steps. Finally, disparity refinement involves a series of calculations for dealing with outliers in the disparity maps.

The literature relevant to the aforementioned steps is described extensively in Chapter 2.

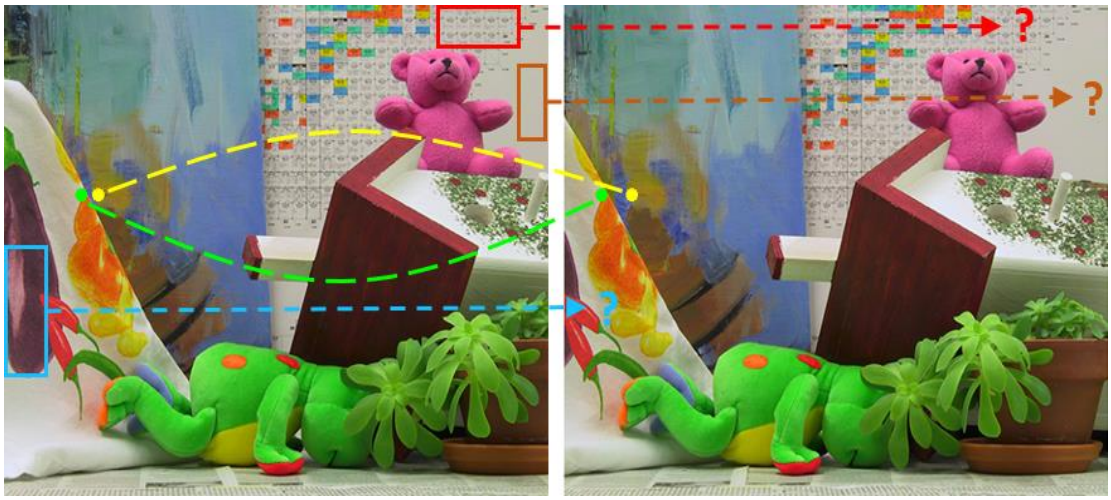


Figure 4. Challenging image areas for disparity estimation.

1.3 Research challenges and objectives

The disparity estimation problem is very challenging, since the correspondence estimation process is hindered due to several factors. First of all, it is difficult to establish correspondences between pixels that lie inside low-textured image areas.

For example, for the pixels that lie inside the brown box on the left image of Figure 4, it is ambiguous which their corresponding pixels on the right image are. Additionally, there are image areas that are visible in one image and invisible in the other image, such as the area of the left image that lies inside the cyan box in Figure 4. It is challenging to deduce the disparity of these areas from neighbouring areas that are visible in both images. Regions near disparity discontinuities may also give inaccurate disparity estimates. The yellow and green pixels in Figure 4 are examples of pixels near depth discontinues. Moreover, there are areas with repeated texture, such as the area inside the red box, which may cause inaccurate correspondences. Another factor that hinders disparity estimation is illumination changes.

Though mature, the estimation of dense disparity maps from stereo image pairs is still a challenging task, since there is sufficient space for improving accuracy, minimizing the computational cost and providing new ways of handling efficiently low-textured areas, outliers and depth discontinuities.

This PhD proposes two methodologies to fulfil the objectives of handling efficiently low-textured areas, disparity discontinuities, repeated textures and outlier areas.

1.4 Contributions of the proposed methodologies

This PhD thesis proposes two methodologies for the accurate estimation of dense disparity maps, which are referred in the following as “methodology A” and “methodology B”. The synoptic contributions of these methodologies are provided in this section.

1.4.1 Contributions of methodology A

Methodology A is the first approach in the literature that combines efficiently colour, CENSUS and SIFT information. In more detail, colour, CENSUS and SIFT information is used to define three matching cost metrics. The matching cost metrics are aggregated using adaptive weights and their cost volumes are acquired. A novel two-phase strategy is then applied to merge the individual cost volumes into a

combined one. The strategy followed for the combination could inspire future approaches to fuse cost volumes, which have been acquired from different types of cost metrics.

Methodology A exploits efficiently image segmentation in the disparity optimization and disparity refinement steps.

Regarding the disparity optimization step (which helps to improve the disparity estimation in low-textured regions and in regions with repetitive texture), image segmentation is used to introduce a new criterion for the definition of the smoothness penalty terms that are used in a semi-global disparity optimization method. This criterion helps to improve the accuracy of the original semi-global method.

Image segmentation is also used for different tasks within the disparity refinement step. One of these tasks is the outliers handling, where methodology A proposes a new approach. Briefly, in this approach inlier pixels of a segment propagate their disparities to the outlier pixels. When, there is not a sufficient number of inlier pixels in a segment, the disparity information from reliable neighboring segments is used to define the disparity of this segment. A second important task, within the disparity refinement step, concerns the handling of large low-textured areas. This task is based on novel disparity histogram analysis, which helps to filter out outlier disparities from large low-textured image regions, before applying disparity plane fitting in each region using the remaining reliable disparities. The exploitation of histogram analysis contributes to improve the accuracy of the disparity plane fitting. Methodology A also introduces a two-step approach to refine the disparity information at the depth discontinuities.

The most computationally expensive parts of methodology A are suitable for Graphics Processing Units implementation (GPU). This fact helps to minimize the total computational cost of methodology A.

1.4.2 Contributions of methodology B

Methodology B combines Gabor, gradient and colour information to define a matching cost metric. This cost metric, which is newly introduced in the literature, can be rapidly estimated, while contributing to the improvement of the accuracy.

Moreover, methodology B provides a novel approach of using guided image filtering for the cost aggregation step. This approach suggests to apply guided image filtering separately for support windows of two different sizes and then to select the appropriate support window size for each pixel based on the texture homogeneity within the local region around this pixel. Therefore, for low-textured areas the larger one support window is preferred in order to contain more discriminative information. This helps to enhance the accuracy of matching, since the matching between regions that contain discriminative information is more reliable. The suggested approach of using guided image filtering contributes to acquire better disparity estimation results than following other literature approaches, which also use guided image filtering.

With respect to the disparity optimization step, methodology B proposes innovative weighted semi-global disparity optimization, where the path costs of a considered pixel may have different weights depending on the pixels that precede the considered pixel along each path direction. The weighted semi-global optimization improves the disparity estimation in image regions with low or repeated textures and gives better results than the original semi-global method [38].

Regarding the disparity refinement step, the handling of the outlier areas is performed by exploiting a straightforward scheme that slightly increases the computational cost. This scheme examines the inlier pixels that lie on the left and the right sides of an outlier pixel before propagating a disparity value to the outlier pixel. Finally, methodology B introduces an efficient technique for correcting disparity errors at depth discontinuities.

The most computationally expensive parts of methodology B can be implemented in GPU, therefore reducing the computational cost of methodology B.

1.4.3 Conclusions on contributions of methodology A and methodology B

To sum up, methodology A and methodology B introduce innovative ideas related to the stereo disparity estimation steps, thus contributing positively to the research objectives of handling outliers, low-textured areas, repeated textures and depth discontinuities. The fulfillment of these objectives helps methodology A and

methodology B to achieve high disparity estimation accuracy. A more complete description of the contributions is given in Chapter 3.

1.4.4 Contributions of the approach developed for dense stereo 3D point cloud generation

Evidently, the main focus of this PhD thesis is to present novel solutions for estimating dense disparity maps form short-baseline stereo image pairs. Nevertheless, during the PhD study, it was invested some research effort to develop an approach that aims to improve the accuracy of the 3D point clouds, which are generated from wide-baseline stereo pairs.

Initially, this approach applies some criteria to select appropriately the stereo image pairs to be used for 3D point cloud generation. Then for a selected stereo pair, the DAISY descriptor is exploited to estimate dense correspondences between the rectified images of the stereo pair. Afterwards, this approach applies a novel two-phase strategy to remove outliers, while the accuracy of the generated 3D point cloud is improved by combining sub-pixel accuracy estimation and the moving least squares algorithm.

This approach contributes to acquire point clouds of better accuracy when compared to the point clouds that are generated using descriptor-based matching in pixel accuracy. Additionally, this approach contributes to remove multiple point cloud outliers.

1.5 Outline

The current thesis is laid out as described below. Chapter 2 contains the literature review relevant to the stereo disparity estimation problem. In specific, the disparity estimation methodology has been divided into generic steps and the background relevant to each step is presented in this chapter. Chapter 2 also provides information on non-conventional disparity estimation techniques that fuse different sensor types.

Chapter 3 gives an overview of the two methodologies developed through the

PhD period and compares those methodologies with literature. Moreover, Chapter 3 describes the preprocessing steps that are applied prior to the methodologies.
In Chapters 4 to 6 the detailed description and the experimental evaluation of the methodologies that were developed during the PhD are provided. In particular, Chapter 4 provides details on the pixel-based cost measures and the cost aggregation approaches that are considered in the presented methodologies. Chapter 5 contains information regarding the disparity optimization step and the disparity refinement approaches for handling problematic areas, which include outlier areas, uniform areas and repeated textures. Chapter 6 provides information on the used parameters and the experimental results. Chapter 7 gives an overview of the work completed so far with respect to the short-baseline stereo disparity estimation and suggests several directions of possible future work. In Appendix A the disparity and error results of the extended stereo dataset are given.

While, Chapters 3 to 7 deal with the short-baseline stereo disparity estimation, which is the main subject of this PhD thesis, in Appendix B a methodology that assists in improving the accuracy of stereo point clouds that are extracted from stereo pairs with wide-baseline is presented.

Chapter 2

2. Stereo vision background

2.1 Generic steps of stereo disparity estimation

The work in [1] presents a complete taxonomy of approaches used for stereo disparity estimation. The categorization of the approaches is based on the following four generic steps, into which most of the stereo algorithms can be decomposed: (a) matching cost computation; (b) cost aggregation; (c) disparity optimization; and (d) disparity refinement. In sections 2.2 and 2.3, the literature review relevant to matching cost computation and cost aggregation is presented, respectively. Section 2.4 describes existing disparity optimization approaches, while section 2.5 reports on disparity refinement techniques. Additionally, sections 2.2 to 2.5 include subsections that briefly mention the literature exploited by the proposed methodologies. Finally, in section 2.6 other non-conventional disparity estimation techniques are described.

2.2 Matching cost computation

Matching computation deals with the definition of pixel-based cost measures. A pixel-based cost measure determines the matching cost between two pixels that lie on different images. Usually, different types of pixel-based cost measures are combined in order to form efficient cost metrics. Prevalent, pixel-based cost measures include the absolute difference of image intensity values [9], gradient-based measures [9], the sampling insensitive cost measure [9], Gabor-feature-based measures [11], and non-parametric transforms such as CENSUS [12].

Disparity estimation approaches use combinations of individual pixel-based cost measures in order to form a final matching cost term that inherits the advantageous characteristics of each measure. In specific, the works in [14], [15], [16] exploit a combination of absolute intensity differences, as well as the Hamming distance of CENSUS transform coefficients. The cost term used in [17], [18] combines

absolute intensity differences and a gradient based measure. The work in [19] uses a combination of CENSUS, colour and gradient based cost measures. The matching cost term used in [11] integrates Gabor, gradient and colour information.

The matching cost values over all pixels and all candidate disparities form the initial cost volume. If the matching cost over a single pixel is used for disparity estimation, the resulting disparity map will be heavily corrupted by noise. In order to reduce matching ambiguity, pixel-based matching costs are locally aggregated in the initial cost volume. In section 2.3 different approaches for performing cost aggregation are described.

2.2.1 Matching cost terms used in the proposed methodologies

Methodology A, following the paradigm of algorithms [14], [15], [16] uses a combination of colour and CENSUS information to define cost terms. However, methodology A does not rely only on colour and CENSUS information, but also studies how SIFT information could be efficiently used to increase the accuracy of the disparity results.

Methodology B, similarly to the method described in [11], uses a combination of Gabor, gradient and colour information to define a cost term. The difference between methodology B and the method in [11] is that methodology B uses the sampling insensitive cost measure [9] to exploit colour information, while the method in [11] uses the sum of absolute difference between the pixel intensity values as a cost measure.

2.3 Cost aggregation approaches

The performance evaluation on different cost aggregation approaches, which was presented in [20], shows that until 2008, Adaptive support weight [21] and Segment-support [22] approaches outperformed the rest of cost aggregation approaches. The adaptive support weight method [21] adjusts the weights based on colour similarity and proximity principles. In the Segment-based approach [22], pixels inside the support window that belong to the same segment as the center pixel are

given a weight equal to 1, while pixels inside the support window, which do not belong the same segment, are given a weight according to their colour similarity to the center pixel.

Cost aggregation methods that build a support window with variable size and/or shape, adaptive to the image content, can also be found in the literature. In [23] a fast method, where an upright cross local support skeleton is adaptively constructed for each anchor pixel, is presented. Then, given the local cross-decision results, a shape adaptive full support region is dynamically constructed by merging horizontal segments of the crosses in the vertical neighbourhood.

In recent years, several approaches perform cost aggregation by filtering the initial cost volume. A prevalent filter used for this scope is the bilateral filter [24]. For instance, the work in [25] proposes a recursive implementation of the bilateral filter, where the computational complexity is linear in both input size and dimensionality. Recently, the use of more efficient filters has been proposed. The edge preserving guided image filter [26] has been exploited in [18] and [27]. While, the constant weighted median filter has been proposed and exploited in [29].

2.3.1 Cost aggregation approaches used in the proposed methodologies

Methodology A uses the adaptive support weight [21] cost aggregation approach. This traditional approach is selected because it gives satisfactory disparity estimation accuracy, while at the same time it is suitable for Graphics Processing Units (GPU) implementation [27].

On the other hand, methodology B performs cost aggregation using the guided image filter [26]. This filter helps to achieve good disparity estimation accuracy at a low computational cost. Moreover, guided image filter can be implemented in GPU [27].

2.4 Disparity optimization approaches

After estimating the aggregated costs for all pixels and all disparities, the next

step is to select the optimum disparity of each pixel that best represents the scene surface.

The simplest approach of estimating the disparity of a pixel is to select the disparity with the lowest associated aggregated matching cost. This approach, which is called the Winner-Takes-All (WTA) approach, is used by the local methods [15], [18], [21], [22], [23], [27], [30] that give emphasis on the matching cost computation and the cost aggregation steps and not to the disparity optimization step.

However, the WTA approach does not contribute to the improvement of disparity estimation in low-textured regions and areas were repetitive features occur. Therefore, in order to improve disparity estimation in these areas various disparity optimization approaches have been proposed. Appropriate disparity optimization approaches include cooperative, global and semi-global methods.

Cooperative methods [33], [34] firstly use colour or grayscale information to segment the input images into meaningful non-overlapping regions. Then, they compute the initial disparity estimate of the image by exploiting a prevalent matching algorithm. Afterwards, a disparity fitting technique is employed to perform the task of disparity refinement for each region. Cooperative algorithms reduce the number of regions by clustering regions in the parameter space of the disparity plane before optimization. The main drawback of cooperative methods is that their iterative nature increases their computational burden.

Global optimization methods initially define a global energy function for the whole image. Once the global energy has been defined, global optimization algorithms attempt to estimate for each pixel of the image the disparity d that minimizes this global function. Graph-cuts [31], [32], [35] and belief propagation [17], [36] are two prevalent approaches for minimizing a global function. Though graph-cuts and belief propagation give accurate depth maps, they have increased computational complexity and memory requirements, while graph cuts are also relatively slow [37].

Another category of disparity optimization methods, includes the semi-global methods [38], [39]. Those approaches attempt to find the global minimum for independent scanlines and not for the complete images. Semi-global approaches provide a good balance between computational complexity and accuracy.

Table 1 summarizes the advantages and the disadvantages of the presented

disparity optimization approaches. Semi-global optimization approaches combine both computational efficiency and accuracy, while the rest of the approaches, though they provide high accuracy, they have increased computational cost, too.

Disparity optimization approaches	Algorithms	Advantages	Disadvantages
Cooperative optimization	Cooperative algorithms [33], [34]	– Good accuracy	– High computational cost
Global optimization	Graph cut algorithms [31], [32], [35]	– Good accuracy	– High computational cost – Memory intensive
	Belief propagation Algorithms [17], [36]	– Good accuracy	– Medium computational cost – Memory intensive
Semi-global optimization	Semi-global algorithms [38], [39]	– Good accuracy – Reduced computational cost	

Table 1. Comparison of disparity optimization approaches.

2.4.1 Optimization approach used in the proposed methodologies

The semi-global method [38] is exploited in both of the proposed methodologies to optimize the disparity estimation results. The reason behind this selection is that the semi-global method has decreased computational complexity, while at the same time it contributes significantly to the improvement of the disparity results. The two methodologies of this thesis propose extensions of the original semi-global method, which assist in enhancing the disparity estimation accuracy.

2.5 Disparity refinement techniques

The disparity results have to be refined, since they are polluted with outliers in occluded areas, uniform areas and depth discontinuities.

The disparity value of an occluded point is usually inferred from the disparities of its neighbouring unoccluded pixels. In the approach of [36], the disparity of an

occluded pixel is set equal to the disparity of the horizontally closest left non-occluded neighbour. In the methodology of [18], the disparity value, which is assigned to the occluded pixel, is set equal to the minimum disparity between the non-occluded pixels that lie on the left and the right sides of the occluded pixel, along the horizontal direction. A bilateral filter is used to smooth the filled regions, in order to deal with horizontal artifacts that are produced after applying this outlier filling scheme. The method presented in [17] uses image segmentation to separate images into segments and then solves the disparity estimation problem by estimating a disparity plane for each estimated segment of the scene. In this method, the disparity of an occluded pixel is deduced from the disparity plane it belongs to. The work in [16] uses iterative region voting to fill outliers, where the filled outliers are marked as 'reliable' pixels and used in the next iteration, so that valid disparity information can gradually propagate into outlier regions.

In [40], two approaches for reliably filling outliers are proposed. The weighted least squares approach is based on absolute colour difference and weighted least squares, while the segmentation-based least squares approach is based on least squares with segmented points. The second approach is more accurate than the first one, but it requires much more computational cost.

2.5.1 Refinement techniques used in the proposed methodologies

Methodology A does not rely on any particular refinement approach to build its disparity refinement approach. However, it bears general similarity to methods, such as [17] and [40], which also utilize image segmentation to perform the disparity refinement task.

Methodology B uses as basis the refinement approach of methodology [18] and builds upon it a straightforward disparity refinement scheme, which yields to better refinement results than the basis approach, with minor increase in the computational cost.

2.6 Non-conventional disparity estimation approaches

Disparity estimation using stereo vision is a cost efficient solution for generating depth maps. However, there are other options to perform this task. In recent years, low-cost sensors, such as Kinect or Time-of-Flight (TOF) sensors, are used to estimate depth maps.

Though low-cost sensors give reliable depth estimates for surfaces with low or repeated textures, their spatial and depth resolution is low. On the other hand, a stereo vision system, which uses high resolution cameras, can provide high resolution depth maps. However, stereo vision systems usually show difficulty in estimating depth in low-textured surfaces and surfaces with repeated textures.

Some recent works attempt to fuse depth sensors with stereo vision systems in order to combine the advantages of both systems.

In [41], a TOF sensor is combined with a stereo vision system. The data term, which is used by the introduced Markov Random Field to estimate depth maps of high accuracy, is formed from a weighted linear combination of a cost function of stereo matching and a cost function of the TOF sensor.

The method in [42] computes a TOF sensor confidence map and a stereo confidence map based on local image features. The two confidence maps are then incorporated into the cost function, which is used to populate the 3D volume created by a plane-sweeping stereo matching algorithm. The final depth map is acquired after applying WTA to this 3D volume.

The work in [43] proposes a global optimization scheme that combines depth information from Kinect with stereo matching. The fusion of both sensors helps to obtain correct scene depth in ambiguous areas and fine structural details in textured areas.

2.7 Summary

Chapter 2 provided the literature related to the generic steps of binocular stereo vision, which are the following: (i) matching cost computation, (ii) cost aggregation, (iii) disparity optimization and (iv) disparity refinement. Additionally, this

chapter briefly referred to the literature exploited by the two methodologies proposed in this thesis. Chapter 2, finally, described several non-conventional disparity estimation techniques which attempt to fuse depth sensors with binocular stereo vision systems.

The following chapter describes the contributions of methodologies A and B and discusses their differences to relevant state-of-the art methods. Moreover, it gives the flowcharts of methodologies A and B as well as their preprocessing steps.

Chapter 3

3. Overview of the proposed methodologies

As it was aforementioned, this thesis presents two methodologies that were developed during the PhD period. The complete contributions and the flowcharts of these methodologies are given in this Chapter. Additionally, these methodologies are compared with the state of the art. Finally, this Chapter describes the preprocessing steps of the methodologies.

3.1 Contributions of developed methodologies

This section concisely describes the contributions of each of the proposed methodologies.

3.1.1 Contributions of methodology A

The most significant contributions of methodology A include the following:

- This methodology acquires a combined cost volume by exploiting three types of cost metrics. The first cost metric combines RGB-CENSUS information, the second one uses only CENSUS information and the third one SIFT information. The cost metrics are aggregated using adaptive weights and their cost volumes are acquired. A reliable two-phase strategy is then followed to merge the individual cost volumes into a combined one.

  This methodology, to the extent of my knowledge, is the first one that combines efficiently RGB, CENSUS and SIFT information.
- This method exploits image segmentation in several stages of this approach. This methodology applies plane fitting just to segments that correspond to large uniform areas and not to all segments. This fact reduces the dependency of methodology A from the result of the disparity plane fitting, which may be of reduced accuracy for small segment areas, due to the decreased number of contained disparities. Also, a metric that verifies if planar fitting is successful is

used, since not all large uniform areas can be considered as planar.

Segmentation is also useful in the disparity optimization step. In more detail, the mean-shift segmentation maps of the stereo pair are used to introduce a new criterion for the definition of the smoothness penalty terms that are used in the original semi-global scanline optimization method of [38] (previously exploited, among other works, in [13], [16], [44], [45]). The modified scanline method is employed for the optimization of the combined cost volume.

Moreover, segmentation is exploited for the outliers handling task, where an efficient strategy that incorporates segmentation-based outliers handling to successfully cope with occluded areas, is presented.

- Handling of large uniform areas is based on disparity histogram analysis, which removes outlier disparities from large uniform regions, before applying disparity plane fitting in each region using the remaining reliable disparities.

Except for the major contributions, some secondary contributions of methodology A are the following:

- A weighted variant of the original CENSUS transform, which improves the disparity accuracy, is proposed.
- Disparity refinement at disparity discontinuities is performed by applying a two-step disparity edges refinement approach. The first step handles disparity errors at depth discontinuities in a coarser level and the second one in a finer level.

3.1.2 Contributions of methodology B

Most significant contributions of methodology B include the following:

- An efficient cost metric for estimating the similarity between two pixels. The cost metric combines gradient difference, Gabor feature difference and a sampling-insensitive dissimilarity measure. It can be rapidly estimated, while it contributes to the accuracy improvement.
- A novel strategy for exploiting guided image filtering [26]. In brief, the guided image filtering is applied separately for support windows of two different sizes

and the appropriate support window size for each pixel is selected based on the texture homogeneity within the local region around this pixel. Texture homogeneity is examined exploiting the mean-shift segmentation maps [46] of the stereo pair.

- An innovative weighted variant of the semi-global optimization method of [38], where the path costs of a considered pixel may have different weights depending on the pixels that precede the considered pixel along each path direction. This feature improves the overall performance of the original semi-global method.
- A novel simple scheme, which assists in successfully handling outliers. This scheme examines if the pixels on the right or on the left side of the outlier pixel are more similar in terms of colour to that pixel, before assigning a disparity value to it. Finally, an efficient technique for correcting disparity errors at depth discontinuities is introduced.

3.2 Proposed methodologies in comparison to state-of-the-art methods

This section gives the differences between the proposed methodologies and state-of-the art methods in this field.

3.2.1 Methodology A in comparison to state-of-the-art methods

Methodology A is the first one that combines RGB, CENSUS and SIFT information by utilizing an efficient strategy. There are several works that use RGB and/or CENSUS information, such as [12], [14], [15], [16], [45], [47], but they do not exploit the SIFT information, which could probably improve their performance. However, the approaches that use SIFT descriptors, or similar ones (such as SURF [48]), for the case of short-baseline stereo disparity estimation, are limited. For instance, the work in [49] combines mutual information, SIFT descriptors and segment based plane-fitting to find correspondences for stereo image pairs which undergo radiometric variations. The paper in [50] uses SURF key points for the initial disparity estimation,

which is further improved by using graph cuts for disparity plane assignment.

Many methods, such as [17], [33], [35], [50], [51], exploit image segmentation algorithms in order to separate images into segments and then solve the disparity estimation problem by assigning, in various ways, a disparity plane for each estimated segment of the scene. In contrast to this class of approaches, the proposed method applies plane fitting only to large segments that correspond to low-textured areas. Additionally, a metric that verifies the success of plane fitting is used to prevent application of plane fitting to low-textured areas that are not (near) planar.

The disparity histogram analysis, described in this thesis, could be used as preprocessing step in algorithms that perform plane fitting using methods that are sensitive to outliers, such as the least square error (LSE) based plane fitting algorithm, which is used in [35] and [51]. Even plane fitting algorithms that are insensitive to outliers, such as RANSAC (Random Sample Consensus) [52], could be fostered by this outlier filtering technique, since their computational cost would be reduced in case the data to be fitted contains less outliers. Disparity estimation methods that exploit RANSAC plane fitting include [36], [50] and [51].

Many methods, such as [11], [16], [17], [22], [23], [33], [44], [53], [54], are evaluated using just the four well-known stereo pairs of the Middlebury Stereo Online Evaluation Benchmark and some of them [16], [17], [33] manage to rank among the top methods. However, there are additional Middlebury stereo pairs that can be used to present a more thorough and complete evaluation. Methodology A, except for the well-known stereo pairs, uses 27 more stereo pairs for assessing the overall performance of this approach.

3.2.2 Methodology B in comparison to state-of-the-art methods

Several methods require iteration cycles in order to improve gradually the accuracy of the estimated disparity maps [33], [53], [54], [55]. Consequently, the number of iterations affect the computational cost of an approach. On the contrary, the proposed method gives disparity results of superior accuracy without performing any repetitive refinement.

Plenty of methods, such as [17], [33], [35], [50], exploit image segmentation

algorithms in order to separate images into segments and then solve the disparity estimation problem by assigning a disparity plane for each estimated segment of the scene. In contrast to this class of approaches, the proposed method does not require plane fitting to give accurate disparity results.

Methodology B, such as methodology A, also uses 27 additional stereo pairs for assessing its overall performance.

3.3 Flowchart of the proposed methodologies

The developed methodologies can be concisely demonstrated using flowcharts, which are presented in this subchapter.

3.3.1 Flowchart of methodology A

Methodology A is divided into four steps, as visualized in the flowchart of Figure 5.

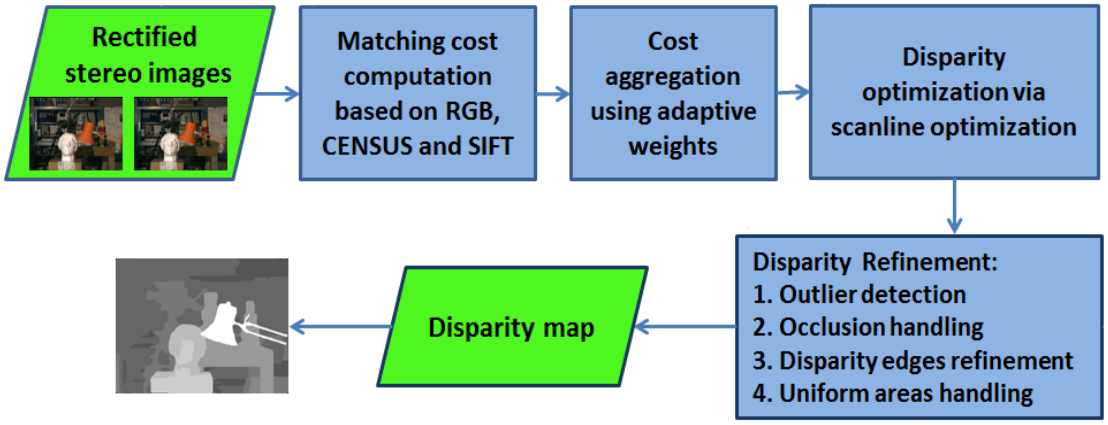


Figure 5. Flowchart of methodology A.

The matching cost computation and cost aggregation steps are described in Chapter 4. While, the disparity optimization and disparity refinements steps are described in Chapter 5.

3.3.2 Flowchart of methodology B

Methodology B is also divided into four steps, as visualized in the flowchart of Figure 6.

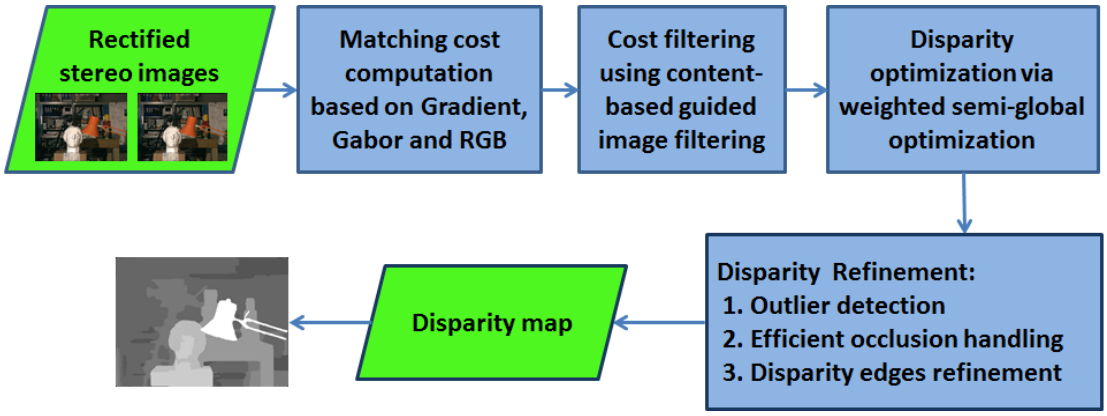


Figure 6. Flowchart of methodology B.

The matching cost computation and cost filtering steps are described in Chapter 4. While, the disparity optimization and disparity refinements steps are described in Chapter 5.

3.4 Preprocessing steps

3.4.1 Rectified image pairs

As it is mentioned in subsection 1.2.1, the input stereo image pair should be rectified, so that the epipolar lines become horizontal [56]. Therefore, the search of point-correspondences between the two images can be performed along the same horizontal epipolar line. Except for limiting searching area, rectified input makes simpler the application of optimization algorithms, such as the scanline optimization used in this work that uses specific path directions. Additionally, since the resulting rectified images have similar scale and the epipolar lines have the same orientation, it is feasible to define and compare adaptive support areas of the same size and orientation between two rectified images. Any efficient existing algorithm, such as the one in [56], can be used for the rectification task.

3.4.2 Radiometric alignment of stereo images

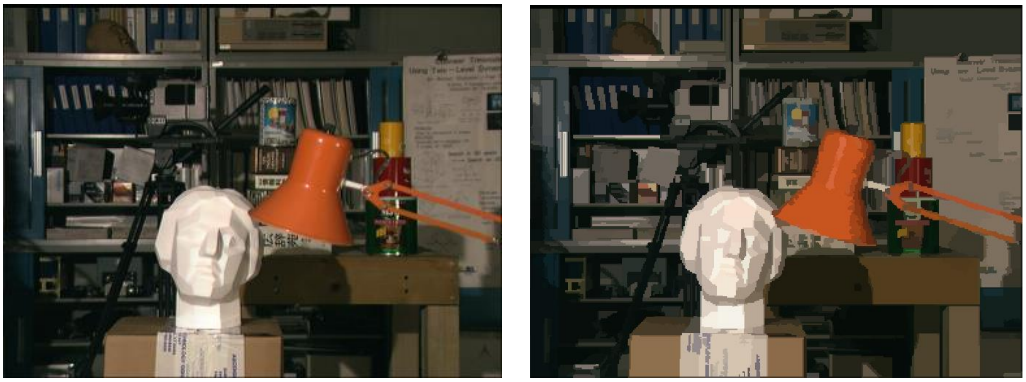
Stereo images may have inconsistent colour values between corresponding pixels due to unknown various radiometric variations. This PhD thesis does not deal with disparity estimation from image pairs that are affected from radiometric variations. However, some approaches that attempt to radiometrically align stereo images are mentioned for completeness.

The work in [57] proposes a radiometric calibration method to align multiple images of moving cameras. This method defines the Brightness Transfer Function through the joint histogram produced by Normalized Cross Correlation based stereo matching, and then it estimates the camera response function and vignetting effects between images. In [58], a colour mapping method, between images acquired under different acquisition conditions, is suggested. This approach uses SIFT features to find a minimal set of piecewise consistent colour mappings assuming planar regions. The method in [59] transforms the input colour images to log-chromaticity colour space from which a linear relationship can be established during constructing a joint probability density function of transformed left and right colour images. From this joint probability density function, a linear function that relates the corresponding pixels in stereo images can be estimated.

3.4.3 Image segmentation

The developed methodologies exploit image segmentation for various tasks. The mean-shift segmentation software (EDISON software [60]), which relies on colour and edge information is used to segment images into non-overlapping regions. Detailed information about the mean-shift segmentation and the EDISON software can be found in [46], [62], [63]. The parameters used for the mean-shift segmentation are the segmentation spatial radius σ_s, which is set to $\sigma_s = 3$ and the segmentation feature space radius σ_r, which is set to $\sigma_r = 3$. The selection of these strict values ensures that the segmentation map will be of high reliability, meaning that most likely a segment will not overlap a depth discontinuity, and this fact is verified also in [22] and [61]. The mean-shift segmentation map for the “Tsukuba” left image (see Figure

7a) is visualized in Figure 7b. The pixels that belong to the same mean-shift segment have an individual label and their mean colour value is computed. Let the labels image be denoted as Lab. The segmentation maps of the left and the right image are computed once and then used in the following algorithmic steps.



(a) **(b)**

Figure 7. Illustration of (a) the left ``Tsukuba'' image and (b) its mean-shift segmentation map.

3.5 Summary

This chapter gave analytically the contributions of methodologies A and B and described their differences to state of the art approaches as well. The flowcharts, which show the outline of the methodologies steps, were also given in the current chapter. Finally, this chapter described the rectification and mean-shift segmentation pre-processing steps and provided information regarding approaches that perform radiometric alignment of stereo images.

The next chapter gives a deep insight into the matching cost computation and the cost aggregation steps of methodology A and methodology B.

Chapter 4

4. Matching cost computation and cost aggregation

4.1 Matching cost computation

This section describes the pixel-based matching costs that are exploited by the proposed methodologies to measure the similarity between two pixels.

4.1.1 Pixel-based matching costs for methodology A

The cost metrics used in methodology A rely on pixel similarity measures that are defined using (i) RGB information, (ii) CENSUS transforms and (iii) SIFT coefficients. This choice is made for the following reasons:

- Exploitation of RGB information gives better results in areas where depth discontinuities exist.
- CENSUS is able to cope with radiometric changes and noise [16].
- The exploitation of SIFT improves the results in textured unoccluded areas, as verified in subsection 4.2.1.2.

4.1.1.1 Weighted CENSUS transform

Methodology A proposes a modification of the original CENSUS transform, which is defined as "weighted CENSUS transform". This modification is described after the definition of the original CENSUS transform.

In order to define the original CENSUS transform [12], a function ξ, which represents the relationship between the intensity of a pixel $\mathbf{x} = (x, y)^{\mathrm{T}}$ and a neighbour pixel $\mathbf{x}_{\mathrm{n}}$, is used:

$$\xi(\mathbf{x}, \mathbf{x}_{\mathrm{n}}) = \begin{cases} 1, \text{if } I(\mathbf{x}_{\mathrm{n}}) < I(\mathbf{x}) \\ 0, \quad \text{otherwise,} \end{cases} \tag{2}$$

where $I(\mathbf{x})$ represents the image intensity of pixel $\mathbf{x}$.

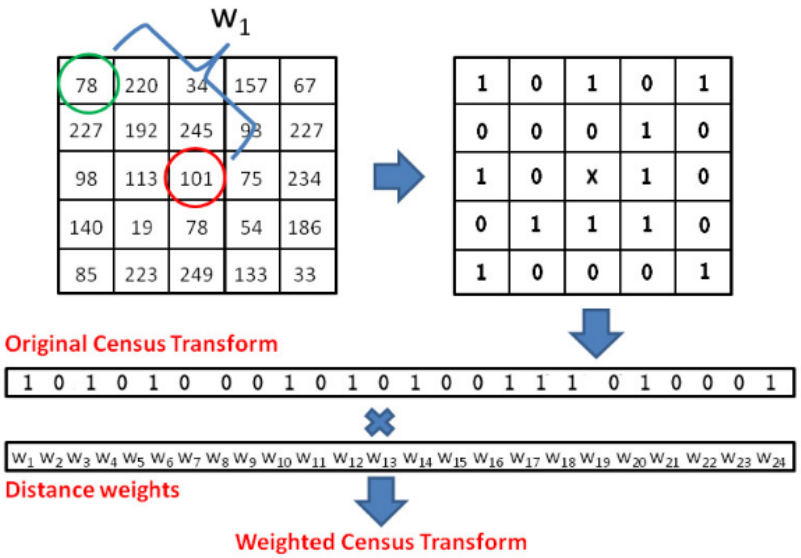


Figure 8. Weighted CENSUS Transform.

The CENSUS transform for pixel $\mathbf{x}$ is computed by comparing its intensity with the intensity of other pixels $\mathbf{x}_n$ that lie within a square window $\mathcal{N}(\mathbf{x})$ around $\mathbf{x}$. The results of these comparisons are then concatenated into a single CENSUS binary vector. Thus, the CENSUS transform of a pixel $\mathbf{x}$ is defined as:

$$CENSUS(\mathbf{x}) = \bigotimes_{\mathbf{x}_n \in \mathcal{N}(\mathbf{x})} \xi(\mathbf{x}, \mathbf{x}_n), \quad (3)$$

where $\otimes$ represents the concatenation operation.

In the proposed "weighted CENSUS transform" (see Figure 8) the bit string that is generated from the original CENSUS transform for a central pixel $\mathbf{x}$, is multiplied by a weight vector, whose elements correspond to the weights between $\mathbf{x}$ and each pixel $\mathbf{x}_n \in \mathcal{N}(\mathbf{x})$. The weight between the central pixel $\mathbf{x}$ (red circle in Figure 8) and a pixel $\mathbf{x}_n$ (green circle in Figure 8) is defined as:

$$\mu(\mathbf{x}, \mathbf{x}_n) = 1 - \beta \cdot \Delta_e(\mathbf{x}, \mathbf{x}_n), \quad (4)$$

where $\Delta_e(\mathbf{x}, \mathbf{x}_n)$ is the Euclidean distance between $\mathbf{x}$ and $\mathbf{x}_n$ and β is a constant parameter. The window size of the weighed CENSUS transform is set experimentally to 5x5. The weight vector gives greater weights for pixels closer to the central pixel, since they are considered as more reliable than those which lie further. Let us denote as $\mathbf{CEN}(\mathbf{x}, c)$ the weighted CENSUS transform at pixel $\mathbf{x}$ for the colour band $c \in \{R, G, B\}$.

4.1.1.2 SIFT-based cost

The SIFT coefficients are extracted densely from an image using the SIFT implementation that was used in the work of [64], which originally deals with visual concept classification. In detail, the parameters used for the SIFT coefficients extraction were selected as: Size of subregions $N_P = 1$, Scale of Gaussian Derivatives $\sigma_{DOG} = 1$, and Number of subregions $N_S = 2$. These parameters define a SIFT descriptor composed of N_S x N_S subregions with subregions' size equal to N_P x N_P pixels. The horizontal and vertical responses for SIFT are calculated using a Gaussian derivative filter, while the diagonal responses are calculated using a fast anisotropic Gaussian derivative filter [65], both using a scale of σ_{DOG}. When, a larger support area was used for the extraction of the SIFT descriptor vector (by increasing N_P and/or N_S), the "foreground fattening" effect [1] was becoming more intense in the estimated disparity map. Let us denote as $\mathbf{SIFT}(\mathbf{x}, \mathrm{c})$ the SIFT descriptor at pixel $\mathbf{x}$ for the colour band $\mathrm{c} \in \{\mathrm{R}, \mathrm{G}, \mathrm{B}\}$.

4.1.1.3 Cost metrics

In this subsection, the way that RGB, weighed CENSUS and SIFT are used to define the similarity between pixels, is described. Given a pixel $\mathbf{x}$ on the left image (reference image) $I_l(\mathbf{x})$, the corresponding pixel on the right image (target image) I_r for a candidate disparity $\mathbf{d}$ will be $I_r(\mathbf{x}^{\mathbf{d}})$, where $\mathbf{x}^{\mathbf{d}} = \mathbf{x} - \mathbf{d}$ and $\mathbf{d} = (d, 0)^{\mathrm{T}}$, since the input stereo images are rectified and consequently the disparity has only a horizontal component. The individual pixel similarity measures $C_{\mathrm{RGB}}(\mathbf{x}, \mathbf{d})$, $C_{\mathrm{CENSUS}}(\mathbf{x}, \mathbf{d})$ and $C_{\mathrm{SIFT}}(\mathbf{x}, \mathbf{d})$ are given from:

$$C_{\mathrm{RGB}}(\mathbf{x}, \mathbf{d}) = \sum_{\mathrm{c} \in \{R,G,B\}} | I_l(\mathbf{x}, \mathrm{c}) - I_r(\mathbf{x}^{\mathbf{d}}, \mathrm{c}) |, \quad (5)$$

$$C_{\mathrm{CENSUS}}(\mathbf{x}, \mathbf{d}) = \sum_{\mathrm{c} \in \{R,G,B\}} \left\| \mathbf{CEN}_l(\mathbf{x}, \mathrm{c}) - \mathbf{CEN}_r(\mathbf{x}^{\mathbf{d}}, \mathrm{c}) \right\|_1, \quad (6)$$

$$C_{\mathrm{SIFT}}(\mathbf{x}, \mathbf{d}) = \sum_{\mathrm{c} \in \{R,G,B\}} \left\| \mathbf{SIFT}_l(\mathbf{x}, \mathrm{c}) - \mathbf{SIFT}_r(\mathbf{x}^{\mathbf{d}}, \mathrm{c}) \right\|_1. \quad (7)$$

Using the aforementioned measures three different matching costs are defined. A RGB-CENSUS combination cost $C_{\text{R-C}}(\mathbf{x},\mathbf{d})$ (following the paradigm of algorithms [14], [15], [16]), a pure weighted CENSUS-based cost $C_{\text{CEN}}(\mathbf{x},\mathbf{d})$ and a SIFT-based cost $C_{\text{S}}(\mathbf{x},\mathbf{d})$, which are given from:

$$C_{\text{R-C}}(\mathbf{x},\mathbf{d}) = \rho\left(C_{\text{RGB}}(\mathbf{x},\mathbf{d}),\lambda_{\text{RGB}}\right) + \rho\left(C_{\text{CENSUS}}(\mathbf{x},\mathbf{d}),\lambda_{\text{CEN}}\right), \quad (8)$$

$$C_{\text{CEN}}(\mathbf{x},\mathbf{d}) = \rho\left(C_{\text{CENSUS}}(\mathbf{x},\mathbf{d}),\lambda_{\text{CEN}}\right), \quad (9)$$

$$C_{\text{S}}(\mathbf{x},\mathbf{d}) = \rho\left(C_{\text{SIFT}}(\mathbf{x},\mathbf{d}),\lambda_{\text{SIFT}}\right), \quad (10)$$

where $\rho(C_y,\lambda_y) = 1 - e^{\left(-C_y/\lambda_y\right)}$.

The exponential function $\rho(C_y,\lambda_y)$ has the advantage of mapping the values of a measure in the range of [0, 1]. This allows different types of measures with different ranges to be scaled into the same range and then to be combined. Additionally, this function allows trimming of outlier values of C_y, depending on the value of λ_y.

Matching cost volumes $C_{\text{R-C}}(\mathbf{x},\mathbf{d})$, $C_{\text{CEN}}(\mathbf{x},\mathbf{d})$ and $C_{\text{S}}(\mathbf{x},\mathbf{d})$ are three-dimensional arrays which store the matching costs for all pixels and all possible disparity candidates. The disparity maps which are acquired after applying WTA to $C_{\text{R-C}}(\mathbf{x},\mathbf{d})$, $C_{\text{CEN}}(\mathbf{x},\mathbf{d})$ and $C_{\text{S}}(\mathbf{x},\mathbf{d})$, respectively, are heavily corrupted by estimation error noise. The severe estimation error noise is obvious in Figure 9, which depicts the disparity map that corresponds to $C_{\text{R-C}}(\mathbf{x},\mathbf{d})$.

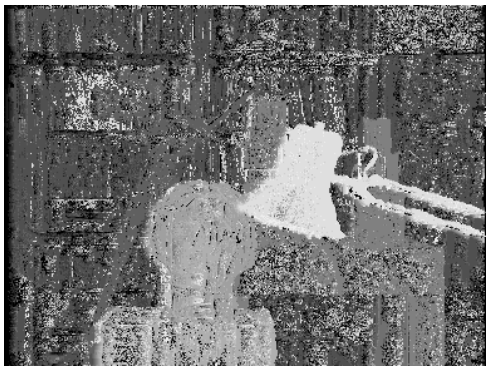


Figure 9. Disparity map after applying WTA to $C_{\text{R-C}}(\mathbf{x},\mathbf{d})$.

4.1.2 Pixel-based matching costs for methodology B

The cost metric used in methodology B is composed of three individual pixel-based cost terms: (i) a gradient-based cost term, (ii) a Gabor-Feature-Image based term [11] and (iii) a Birchfield-Tomasi dissimilarity term [9]. The reasons for using those three terms to compute a combined cost metric are the following:

- The gradient-based cost term shows high robustness to illumination changes, has strong local minima and can be estimated very fast [9].
- The Gabor-Feature-Image, according to [11] is appropriate for texture representation and discrimination, robust to illumination changes, insensitive to image noise and can be calculated quite fast.
- The Birchfield-Tomasi dissimilarity measure, presented in [9], is insensitive to image sampling.

Let I_l^{c} and I_r^{c} be the left and right colour images of the stereo pair, while I_l and I_r are their respective grayscale images. Given a pixel $\mathbf{x}$ on the left image (reference image), the corresponding pixel on the right image (target image), for a candidate disparity value $\mathbf{d}$ is denoted as $\mathbf{x}^{\mathbf{d}}$.

The gradient-based cost term for a pixel $\mathbf{x}$ and disparity $\mathbf{d}$ is given by:

$$C_{\mathrm{gra}}(\mathbf{x},\mathbf{d}) = \left| \nabla_{\mathrm{H}}(I_l(\mathbf{x})) - \nabla_{\mathrm{H}}(I_r(\mathbf{x}^{\mathbf{d}})) \right|, \tag{11}$$

where $\nabla_{\mathrm{H}}(I(\mathbf{x}))$ denotes the gradient in horizontal direction at pixel on grayscale image I.

The second term, as in [11], is based on the Gabor-Feature-Image, which is generated after applying a Gabor filter on an image. The kernel of the Gabor filter can be expressed as:

$$G(x, y, \lambda, \theta, \sigma, \gamma) = \exp\left(-\frac{x'^2 + \gamma^2 y'^2}{2\sigma^2} \right) \cdot \cos\left(2\pi \frac{x'}{\lambda} \right) \tag{12}$$

where

$$\begin{cases} x' = x \cdot \cos(\theta) + y \cdot \sin(\theta) \\ y' = -x \cdot \sin(\theta) + y \cdot \cos(\theta) \end{cases} \tag{13}$$

In Equation (12) and Equation (13), λ represents the wavelength of the

sinusoidal factor, θ represents the orientation of the normal to the parallel stripes of a Gabor function, σ is the standard deviation of the Gaussian envelope and γ is the spatial aspect ratio. The aforementioned values are set as in [11]: $\{\lambda, \theta, \sigma, \gamma\} = \{3, 3\pi/2, 1.5, 1\}$.

Let $G_{\mathrm{H}}(I_l(\mathbf{x}))$ and $G_{\mathrm{H}}(I_r(\mathbf{x}^{\mathbf{d}}))$ denote the outputs of above vertically-varying Gabor kernel for I_l and I_r, respectively. The cost term $C_{\mathrm{gab}}(\mathbf{x}, \mathbf{d})$ for pixel $\mathbf{x}$ at disparity $\mathbf{d}$ is given by:

$$C_{\mathrm{gab}}(\mathbf{x}, \mathbf{d}) = | G_{\mathrm{H}}(I_l(\mathbf{x})) - G_{\mathrm{H}}(I_r(\mathbf{x}^{\mathbf{d}})) | . \quad (14)$$

The third term is given by:

$$C_{\mathrm{BT}}(\mathbf{x}, \mathbf{d}) = \sum_{\mathrm{c=R,G,B}} \frac{D^{\mathrm{c}}(\mathbf{x}, \mathbf{x}^{\mathbf{d}})}{3}, \quad (15)$$

where $D^{\mathrm{c}}(\mathbf{x}, \mathbf{x}^{\mathbf{d}})$, which is the Birchfield-Tomasi (BT) dissimilarity measure between pixels $\mathbf{x}$ and $\mathbf{x}^{\mathbf{d}}$ [9], is estimated as follows. Initially, the intensities in I_l^{c} and I_r^{c} are interpolated using either a previous or a subsequent pixel along the epipolar line. For instance $I_l^{\mathrm{c}}\left(\mathbf{x} - 1/2\right) = 1/2\left(I_l^{\mathrm{c}}\left(\mathbf{x}-1\right) + I_l^{\mathrm{c}}\left(\mathbf{x}\right)\right)$ is the interpolated intensity value at pixel $\mathbf{x}$ with respect to its previous pixel. Let $\hat{I}_l^{\mathrm{c}}(\mathbf{x}) = \left\{I_l^{\mathrm{c}}\left(\mathbf{x} - 1/2\right), I_l^{\mathrm{c}}\left(\mathbf{x}\right), I_l^{\mathrm{c}}\left(\mathbf{x} + 1/2\right)\right\}$ be the set containing the intensity at pixel $\mathbf{x}$ in image I_l^{c} and the interpolated intensities with its previous and subsequent pixel. In the same manner, $\hat{I}_r^{\mathrm{c}}(\mathbf{x}^{\mathbf{d}})$ is the set containing the intensity at pixel $\mathbf{x}^{\mathbf{d}}$ in image I_r^{c} and the interpolated intensities with its previous and subsequent pixel. The Birchfield-Tomasi dissimilarity measure is estimated by:

$$D^{\mathrm{c}}(\mathbf{x}, \mathbf{x}^{\mathbf{d}}) = \min(a, b), \quad (16)$$

where

$$\begin{cases} a = \max\left\{0, I_l^{\mathrm{c}}(\mathbf{x}) - \max\left(\hat{I}_r^{\mathrm{c}}(\mathbf{x}^{\mathbf{d}})\right), \min\left(\hat{I}_r^{\mathrm{c}}(\mathbf{x}^{\mathbf{d}})\right) - I_l^{\mathrm{c}}(\mathbf{x})\right\} \\ b = \max\left\{0, I_r^{\mathrm{c}}(\mathbf{x}^{\mathbf{d}}) - \max\left(\hat{I}_l^{\mathrm{c}}(\mathbf{x})\right), \min\left(\hat{I}_l^{\mathrm{c}}(\mathbf{x})\right) - I_r^{\mathrm{c}}(\mathbf{x}^{\mathbf{d}})\right\} \end{cases} \quad (17)$$

The combined matching cost term, merging Equation (11), Equation (14) and

Equation (15), is expressed as:

$$C(\mathbf{x},\mathbf{d}) = \alpha_1 \cdot \min\left(C_{\text{gra}}(\mathbf{x},\mathbf{d}), T_{\text{gra}}\right) + \alpha_2 \cdot \min\left(C_{\text{gab}}(\mathbf{x},\mathbf{d}), T_{\text{gab}}\right) + \left(1-\alpha_1-\alpha_2\right) \cdot \min\left(C_{\text{BT}}(\mathbf{x},\mathbf{d}), T_{\text{BT}}\right), \quad (18)$$

where α_1, α_2 are balance weights and T_{gra}, T_{gab}, T_{BT} are truncation thresholds.

Cost volume $C(\mathbf{x},\mathbf{d})$ stores the matching costs for all pixels and all possible disparity candidates. The disparity map of Figure 10 is acquired after applying WTA to $C(\mathbf{x},\mathbf{d})$. Evidently, the disparity map of Figure 10 is heavily corrupted by estimation-error noise.

Figure 10. Disparity map after applying WTA to $C(\mathbf{x},\mathbf{d})$.

4.2 Cost aggregation

The disparity maps, which are generated relying only on pixel-based matching cost, are heavily corrupted by disparity noise. In order to improve matching robustness, disparity estimation methods use supporting pixel areas (i.e. pixel neighbourhoods) to aggregate pixel-based matching costs.

4.2.1 Cost aggregation for methodology A using adaptive weights

4.2.1.1 Adaptive cost aggregation

The pixel-based matching costs $C_{\text{R-C}}(\mathbf{x},\mathbf{d})$, $C_{\text{CEN}}(\mathbf{x},\mathbf{d})$ and $C_{\text{S}}(\mathbf{x},\mathbf{d})$ (their estimation is described in subsection 4.1.1) are aggregated spatially over support

regions around each pixel. According to the evaluation studies of [20], [66] the adaptive weight approach [21] produces reasonably accurate disparity maps. Thus, this aggregation approach with slight modifications is used in this work.

More specifically, adaptive support-weight based aggregation applies weights to each of the pixels surrounding the pixel of interest. The adaptive-support weights notion is based on the Gestalt principles of similarity and proximity [21]. The similarity principle assumes that the more similar colour a surrounding pixel has to the central pixel of interest, the more likely it is to belong to the same surface, while the proximity principle assumes that the closer a surrounding pixel is to the central pixel of interest, the more likely it is to belong to the same surface.

In order to describe the adaptive-supports weights notion with mathematical expressions, a pixel of interest $\mathbf{x}$ and a neighbour pixel $\mathbf{x}_n$ are considered. The adaptive weight between $\mathbf{x}$ and $\mathbf{x}_n$, is given by:

$$w(\mathbf{x},\mathbf{x}_n) = e^{\left(\frac{-\Delta I(\mathbf{x},\mathbf{x}_n)}{\gamma_c}\right)} \cdot e^{\left(\frac{-\Delta e(\mathbf{x},\mathbf{x}_n)}{\gamma_e}\right)}, \quad (19)$$

where γ_c and γ_e are constant parameters, $\Delta e(\mathbf{x},\mathbf{x}_n)$ is the Euclidean distance between $\mathbf{x}$ and $\mathbf{x}_n$ and $\Delta I(\mathbf{x},\mathbf{x}_n)$ is given by:

$$\Delta I(\mathbf{x},\mathbf{x}_n) = \sqrt{\sum_{c\in\{R,G,B\}} |I(\mathbf{x},c) - I(\mathbf{x}_n,c)|^2}. \quad (20)$$

Similar to [23], the adaptive weights are computed on the input stereo images after applying a median filter that uses a 2x2 neighbourhood in order to alleviate the impact of image noise and subtle non-Lambertian effects.

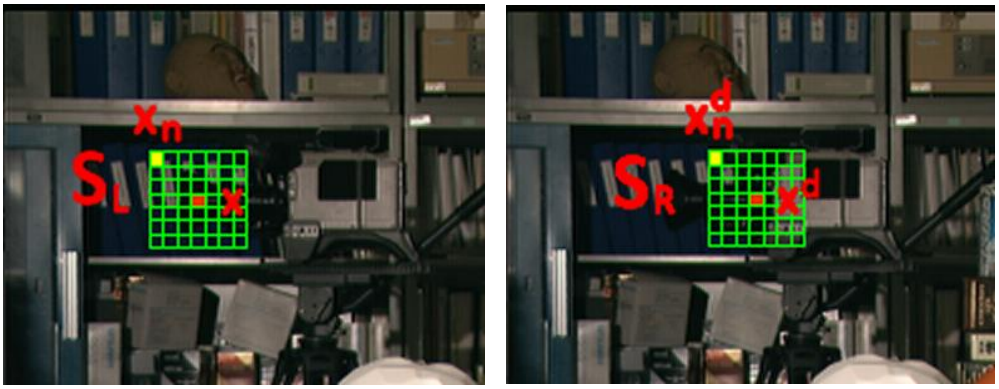


Figure 11. Adaptive weights support region on reference and target "Tsukuba" images.

The adaptive weight approach used in this work has two slight modifications compared to the original work of [21]. Experimental results in [22] proved that the use of the RGB colour space for computing colour similarity decreases the possibility that pixels belonging to different depths are being aggregated in the same support region. For this reason, the RGB colour space is used to compute colour similarity in methodology A, contrary to [21] that uses the CIELab colour space. Additionally, instead of using all pixels in the square support region, only pixels within radius R_S from the central pixel are used. In this way, the support region becomes symmetric around the central pixel $\mathbf{x}$ of interest.

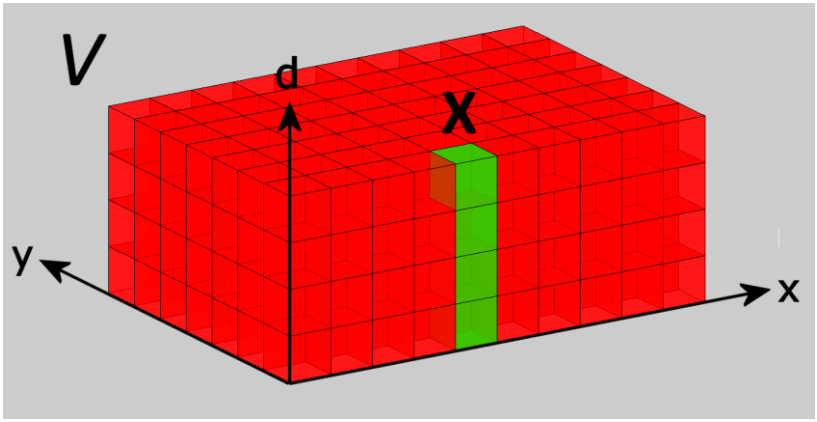
A weight support mask is generated for a pixel $\mathbf{x}$ on the left stereo image, denoted as $w_l(\mathbf{x},\mathbf{x}_n)$. Similarly, a weight support mask is generated for the right stereo image around the corresponding pixel $\mathbf{x}^d$ and is denoted as $w_r(\mathbf{x}^d,\mathbf{x}_n^d)$. Both $w_l(\mathbf{x},\mathbf{x}_n)$ and $w_r(\mathbf{x}^d,\mathbf{x}_n^d)$ are taken into consideration to define the aggregated cost $V(\mathbf{x},\mathbf{d})$ between $\mathbf{x}$ and $\mathbf{x}^d$ as:

$$V(\mathbf{x},\mathbf{d}) = \frac{\sum_{\mathbf{x}_n \in S_L, \mathbf{x}_n^d \in S_R} w_l(\mathbf{x},\mathbf{x}_n)\cdot w_r(\mathbf{x}^d,\mathbf{x}_n^d)\cdot C(\mathbf{x}_n,\mathbf{d})}{\sum_{\mathbf{x}_n \in S_L, \mathbf{x}_n^d \in S_R} w_l(\mathbf{x},\mathbf{x}_n)\cdot w_r(\mathbf{x}^d,\mathbf{x}_n^d)}, \qquad (21)$$

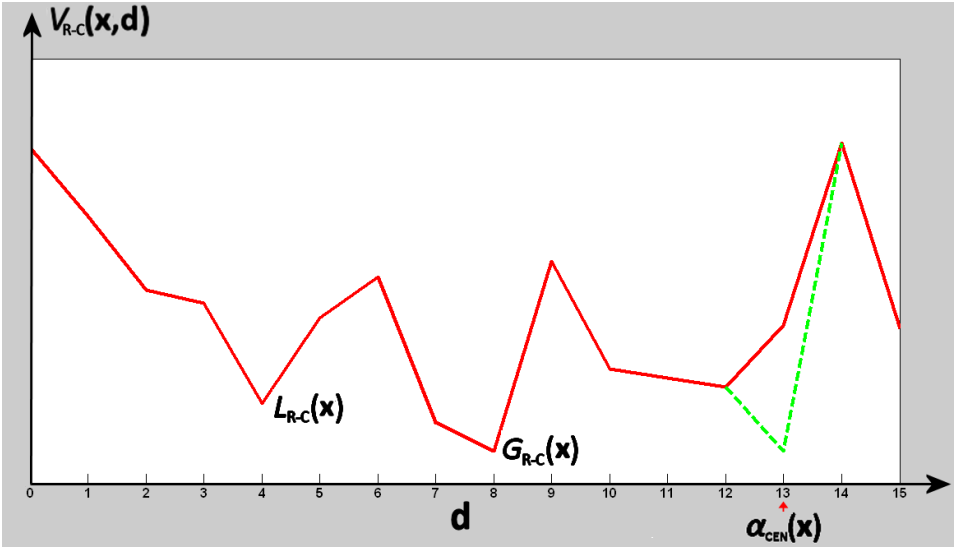
where S_L defines the support region around pixel $\mathbf{x}$ on the left image and S_R the support region around pixel $\mathbf{x}^d$, on the right image, as it is visualized in Figure 11. If cost $C(\mathbf{x},\mathbf{d})$ is replaced by $C_{R\text{-}C}(\mathbf{x},\mathbf{d})$, $C_{CEN}(\mathbf{x},\mathbf{d})$ or $C_S(\mathbf{x},\mathbf{d})$, the aggregated cost volumes $V_{R\text{-}C}(\mathbf{x},\mathbf{d})$, $V_{CEN}(\mathbf{x},\mathbf{d})$ and $V_{SIFT}(\mathbf{x},\mathbf{d})$ can be estimated, respectively. The schematic representation of a cost volume is depicted in Figure 12a.

4.2.1.2 Combination of aggregated cost volumes

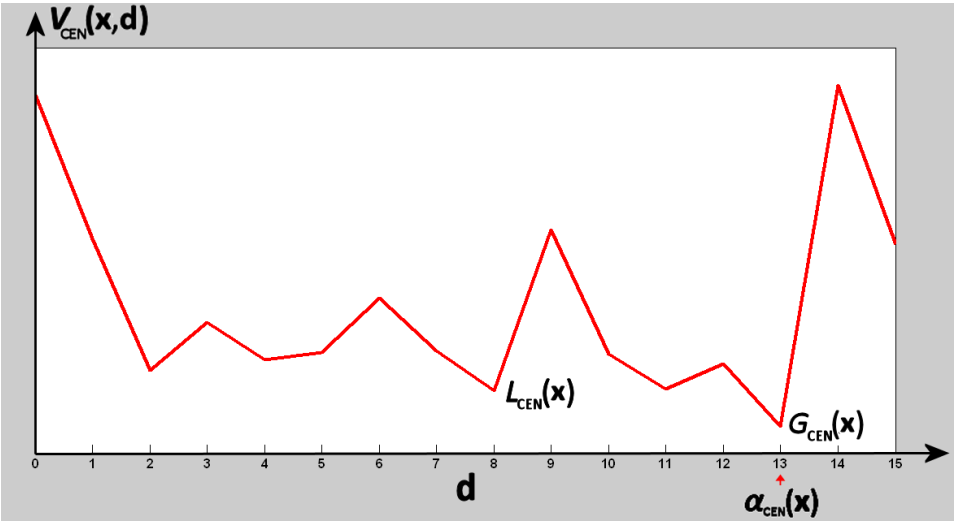
In subsection 4.1.1, the reasons for using the specified cost metrics $C_{R\text{-}C}(\mathbf{x},\mathbf{d})$, $C_{CEN}(\mathbf{x},\mathbf{d})$ and $C_S(\mathbf{x},\mathbf{d})$ are explained. In this paragraph, the details of combining their corresponding cost volumes $V_{R\text{-}C}(\mathbf{x},\mathbf{d})$, $V_{CEN}(\mathbf{x},\mathbf{d})$ and $V_{SIFT}(\mathbf{x},\mathbf{d})$ to produce a "combined" cost volume, are described.



(a)



(b)



(c)

Figure 12. Visualization of (a) a cost volume. The cost variation along disparity for a pixel $\mathbf{x}$ of (b) $V_{\text{R-C}}$ and (c) V_{CEN}.

The proposed approach uses a combination of RGB and CENSUS information via Equation (8) in order to compute $V_{\text{R-C}}$. However, after extensive experiments, it was deduced that the cost volume V_{CEN} computed using only weighted CENSUS information, could be efficiently exploited to refine $V_{\text{R-C}}$. Additionally, it was noticed that the WTA estimated disparity map from V_{SIFT} is reliable for unoccluded textured areas. This fact is exploited here to further refine $V_{\text{R-C}}$. The reason for not combining directly the SIFT information with the RGB and CENSUS information (for instance using an equation similar to Equation (8), with an additional term for SIFT information), is that the ability of the SIFT-based metric to provide accurate disparity estimates at unoccluded textured areas degrades significantly when SIFT is combined directly with other cost metrics, as experimentally verified. In the following, an efficient two-phase strategy for combining $V_{\text{R-C}}$, V_{CEN} and V_{SIFT} is described. This strategy is built upon the aforementioned conclusions regarding V_{CEN} and V_{SIFT}.

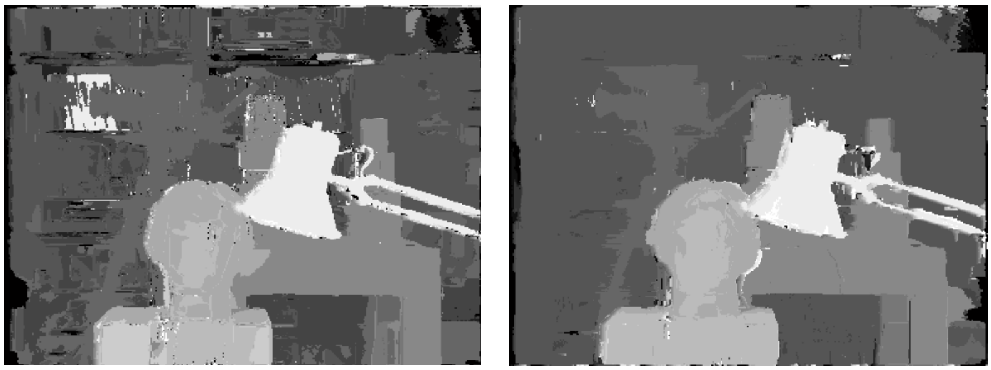
First Combination Phase

During the first phase, $V_{\text{CEN}}(\mathbf{x},\mathbf{d})$ is used to refine $V_{\text{R-C}}(\mathbf{x},\mathbf{d})$. The Peak Ratio confidence measure, one of the best confidence measures according to [67], is used for this purpose.

Peak Ratio confidence measure: Let us consider the curve of cost variation along disparity $\mathbf{d}$ for a pixel $\mathbf{x}$ from cost volume $V(\mathbf{x},\mathbf{d})$. This term is depicted visually with green colour in the visual representation of a cost volume $V(\mathbf{x},\mathbf{d})$ in Figure 12a and an example of cost variation curve is shown in Figure 12c. Let us define as $G(\mathbf{x}) = \min_{\mathbf{d}}\{V(\mathbf{x},\mathbf{d})\}$ the global minimum of $V(\mathbf{x},\mathbf{d})$ and as $L(\mathbf{x})$ the second local minimum of $V(\mathbf{x},\mathbf{d})$. Then, the peak ratio confidence measure is defined as:

$$R(\mathbf{x}) = \frac{L(\mathbf{x})}{G(\mathbf{x})}. \quad (22)$$

Finally, let $\alpha(\mathbf{x}) = \arg\min_{\mathbf{d}}\{V(\mathbf{x},\mathbf{d})\}$, be the optimum disparity value that gives the global minimum of $V(\mathbf{x},\mathbf{d})$. The higher $R(\mathbf{x})$, the more reliable the global minimum of $V(\mathbf{x},\mathbf{d})$ is.



(a) **(b)**

Figure 13. Disparity maps after applying WTA to (a) $V'_{\text{R-C}}$ and (b) V_{SIFT}.

Based on this confidence measure, the optimum disparity for a pixel $\mathbf{x}$, as estimated from $V_{\text{CEN}}(\mathbf{x},\mathbf{d})$, will be "propagated" to $V_{\text{R-C}}(\mathbf{x},\mathbf{d})$. The curves of cost variation along disparity for $V_{\text{R-C}}(\mathbf{x},\mathbf{d})$ and $V_{\text{CEN}}(\mathbf{x},\mathbf{d})$ are depicted in Figure 12b and Figure 12c, respectively.

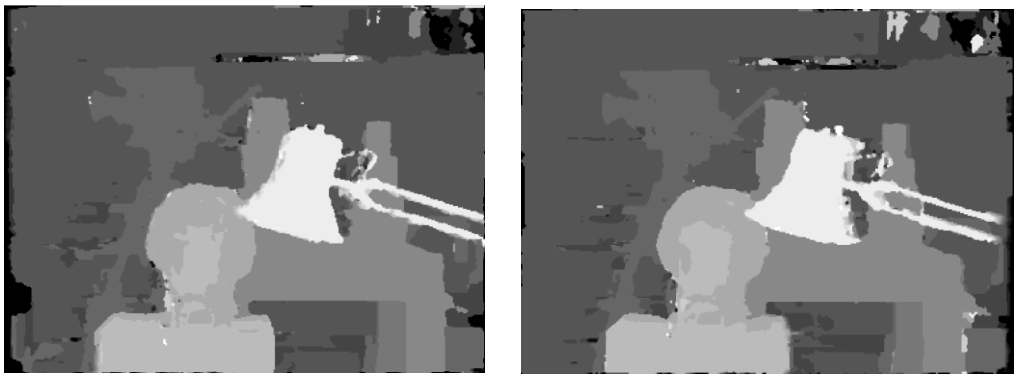
In more detail, for a pixel $\mathbf{x}$, the confidence $R_{\text{R-C}}(\mathbf{x}) = \dfrac{G_{\text{R-C}}(\mathbf{x})}{L_{\text{R-C}}(\mathbf{x})}$ based on $V_{\text{R-C}}(\mathbf{x},\mathbf{d})$ (Figure 12b), is estimated. Similarly, the confidence $R_{\text{CEN}}(\mathbf{x}) = \dfrac{G_{\text{CEN}}(\mathbf{x})}{L_{\text{CEN}}(\mathbf{x})}$ based on $V_{\text{CEN}}(\mathbf{x},\mathbf{d})$ (Figure 12c), is estimated.

In case that $R_{\text{CEN}}(\mathbf{x}) > R_{\text{R-C}}(\mathbf{x})$, at the disparity position $\alpha_{\text{CEN}}(\mathbf{x})$ (position of the global minimum of $V_{\text{CEN}}(\mathbf{x},\mathbf{d})$), the corresponding value of $V_{\text{R-C}}(\mathbf{x},\alpha_{\text{CEN}}(\mathbf{x}))$ is modified according to:

$$V_{\text{R-C}}(\mathbf{x},\alpha_{\text{CEN}}(\mathbf{x})) \leftarrow \min_{\mathbf{d}}\left\{V_{\text{R-C}}(\mathbf{x},\mathbf{d})\right\} - \delta \qquad (23)$$

with $\delta \to 0^{+}$, so that the global minimum of $V_{\text{R-C}}(\mathbf{x},\mathbf{d})$ to coincide with the one of $V_{\text{CEN}}(\mathbf{x},\mathbf{d})$. The part of the curve that changes after this step is depicted with green colour in Figure 12b. The case $R_{\text{CEN}}(\mathbf{x}) > R_{\text{R-C}}(\mathbf{x})$ means that the global minimum of $R_{\text{CEN}}(\mathbf{x}) > R_{\text{R-C}}(\mathbf{x})$ is more confident than that of $V_{\text{R-C}}(\mathbf{x},\mathbf{d})$. In this case, information about the disparity that gives this global minimum is propagated to $V_{\text{R-C}}(\mathbf{x},\mathbf{d})$. After

executing the first phase, $V'_{\text{R-C}}(\mathbf{x},\mathbf{d})$ is acquired. The WTA of $V'_{\text{R-C}}(\mathbf{x},\mathbf{d})$ gives the disparity map of Figure 13a.



(a) **(b)**

Figure 14. Disparity maps (a) $d_{\text{LR}}(\mathbf{x})$ and (b) $d_{\text{RL}}(\mathbf{x})$ after applying second combination phase.

Second Combination Phase

In a second phase, $V_{\text{SIFT}}(\mathbf{x},\mathbf{d})$ is used to refine $V'_{\text{R-C}}(\mathbf{x},\mathbf{d})$. The WTA of $V_{\text{SIFT}}(\mathbf{x},\mathbf{d})$ gives the SIFT-based disparity map $d_{\text{SIFT}}(\mathbf{x})$ (see Figure 13b), which provides reliable disparities in textured unoccluded areas where depth does not change. This is evident in Figure 13b for the disparity of the left ``Tsukuba'' image.

Detection of reliable disparities: In order to find the regions in $d_{\text{SIFT}}(\mathbf{x})$ that are reliable, the mean-shift colour segmentation map (see subsection 3.4.3) is used. If $n(S)$ denotes the number of pixels in a colour segment S and $n_f(S)$ is the number of pixels that have the most frequent disparity in this segment according to d_{SIFT}, then $P_{\mathbf{x}}(S)=\dfrac{n_f(S)}{n(S)}$ adaptive threshold is defined. If $P_{\mathbf{x}}(S)\geq 90\%$, then it is assumed that the disparities inside this segment are reliable (since the vast majority of pixels have the same disparity value).

According to the above, reliable disparities in $d_{\text{SIFT}}(\mathbf{x})$ are propagated to $V'_{\text{R-C}}(\mathbf{x},\mathbf{d})$ in the following way: For every pixel $\mathbf{x}\in S$, the disparity estimate $d_{\text{SIFT}}(\mathbf{x})$ is propagated to $V'_{\text{R-C}}(\mathbf{x},\mathbf{d})$ according to:

$$V'_{\text{R-C}}(\mathbf{x},d_{\text{SIFT}}(\mathbf{x}))\leftarrow \min_{\mathbf{d}}\left\{V'_{\text{R-C}}(\mathbf{x},\mathbf{d})\right\}-\delta \qquad (24)$$

with $\delta \to 0^+$. After executing this second phase, $C_f(\mathbf{x},\mathbf{d})$ is acquired. Let the WTA-estimated disparity map from $C_f(\mathbf{x},\mathbf{d})$ be $d_{\mathrm{LR}}(\mathbf{x})$. After applying a 3x3 median filter on $d_{\mathrm{LR}}(\mathbf{x})$, in order to remove spurious disparities, the disparity map of Figure 14a is generated. By comparing Figure 13a and Figure 14a, it is evident that $d_{\mathrm{LR}}(\mathbf{x})$ can be exploited to efficiently enhance the results in unoccluded textured regions.

Except for the visual demonstration of using the two-phase combination strategy to improve the generated disparity map, an additional numeric evaluation is presented in subsection 6.5.1.2.

If the right image is considered as reference image, then the disparity map $d_{\mathrm{LR}}(\mathbf{x})$ of Figure 14b is acquired.

4.2.2 Cost aggregation for methodology B using guided image filtering

4.2.2.1 Guided image filtering

The matching costs $C(\mathbf{x},\mathbf{d})$ (their estimation is described in subsection 4.1.2) are filtered using the guided image filter [18]. In detail, the filtered cost value of pixel $\mathbf{x}$ at a fixed disparity $\mathbf{d}$ is given by:

$$C'(\mathbf{x},\mathbf{d}) = \sum_{\mathbf{q}} W(\mathbf{x},\mathbf{q}) C(\mathbf{x},\mathbf{d}), \qquad (25)$$

where the filter weights $W(\mathbf{x},\mathbf{q})$ depend on the colour guidance image I (which is the reference stereo image) and they are given from [18]:

$$W(\mathbf{x},\mathbf{q}) = \frac{1}{|w_{\mathbf{k}}|^2} \sum_{(\mathbf{x},\mathbf{q})\in w_{\mathbf{k}}} \left(1 + (I(\mathbf{x}) - \mu_{\mathbf{k}})^T \left(\Sigma_{\mathbf{k}} + \varepsilon U\right)^{-1} (I(\mathbf{q}) - \mu_{\mathbf{k}})\right), \qquad (26)$$

where $|w_{\mathbf{k}}|$ is the total number of pixels in a support window $w_{\mathbf{k}}$ centered at pixel $\mathbf{k}$ and ε is a smoothness parameter. $\Sigma_{\mathbf{k}}$ and $\mu_{\mathbf{k}}$ are the covariance and the mean of pixels colours within $|w_{\mathbf{k}}|$. $I(\mathbf{x})$, $I(\mathbf{q})$ and $\mu_{\mathbf{k}}$ are 3×1 (colour) vectors, while $\Sigma_{\mathbf{k}}$ and the unary matrix U are of size 3×3.

The selection of the appropriate support window size for each pixel, based on its local image content, is discussed in the following paragraph.

4.2.2.2 Selection of the window size based on local image content

This subsection proposes a novel scheme for exploiting guided image filtering. First of all, the shape of the support window is selected to be rectangular and the largest dimension of the support window to be the horizontal one (width). The window's width is twice its height. A support window elongated along the horizontal dimension, i.e., along the dimension in which disparity varies, is used in order to increase the discriminating ability of the window. This fact is experimentally verified in subsection 6.5.2.2.

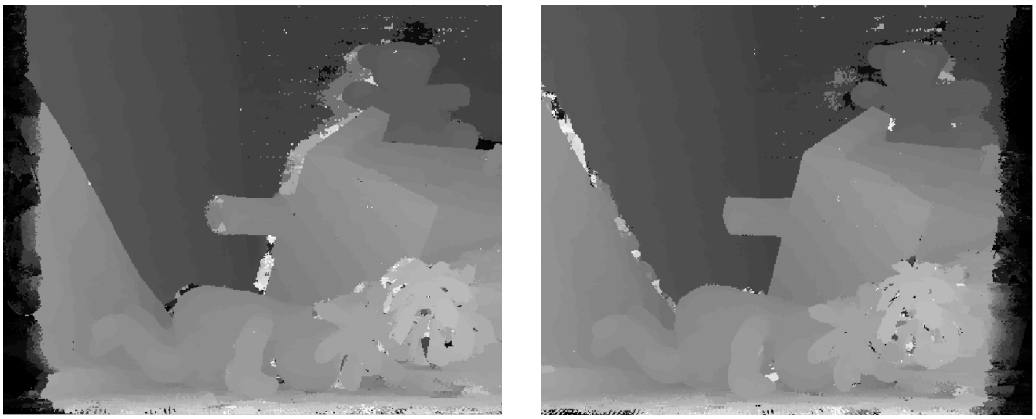
Except for the rectangular shape, in methodology B, windows of two sizes are used. The small window size is R_S x $\lceil R_S / 2 \rceil$ and the large one is $2R_S$ x R_S. The guided image cost filtering is performed separately for both window sizes. Given the two filtered costs, which were estimated by applying guided image filtering for both window sizes, the filtered cost that is finally assigned to a pixel, depends on the local image content around this pixel. In specific, the preferred support window size for each pixel is selected according to the information about the texture homogeneity within the local region around the pixel. Hence, if the neighbourhood around a pixel is homogeneous, then the large support window size, which contains more information, shall be preferred. The criterion to decide which support window is appropriate for a pixel depends on image's segmentation map that provides information about the homogeneity of the image region around a pixel.

In detail, based on image's segmentation map, for each pixel $\mathbf{x}$ the lengths of the ``arms'' stretching to left (F_l), right (F_r), up (F_u) and down (F_d) directions are estimated as visualized in Figure 15a: Given a pixel $\mathbf{x}$ and a direction F_i, $i \in (l,r,u,d)$, the length of $\mathbf{x}$'s arm along the considered direction is given by the number of pixels between $\mathbf{x}$ and the end of the segment where $\mathbf{x}$ belongs. The length is denoted as $\mathrm{M}_i(\mathbf{x})$, $i \in (l,r,u,d)$. The average length of the arms is given by:

$$\overline{\mathrm{M}}(\mathbf{x}) = \left(\mathrm{M}_l(\mathbf{x}) + \mathrm{M}_r(\mathbf{x}) + \mathrm{M}_u(\mathbf{x}) + \mathrm{M}_d(\mathbf{x})\right)/4 \qquad (27)$$

(a) **(b)**

Figure 15. Illustration of: (a) the arms lengths for a pixel $\mathbf{x}$ on a segmentation map and (b) the pixels with support region of $2R_S$ x R_S.



(a) **(b)**

Figure 16. Disparity maps (a) $d_{LR}(\mathbf{x})$ and (b) $d_{RL}(\mathbf{x})$ after applying content based guided image filtering.

If $\overline{M}(\mathbf{x}) > R_S$ then it is assumed that $\mathbf{x}$ lies inside a homogeneous area. Hence, the large window of size $2R_S$ x R_S is considered as the appropriate support window for $\mathbf{x}$, in order to contain more information. For pixels with $\overline{M}(\mathbf{x}) \leq R_S$, the appropriate support window is the one with size R_S x $\lceil R_S / 2 \rceil$. In Figure 15b the pixels for which the appropriate support window has size $2R_S$ x R_S are visualized with red.

Let the filtered cost using the small window be denoted as $C_1'(\mathbf{x},\mathbf{d})$, while $C_2'(\mathbf{x},\mathbf{d})$ denotes the filtered cost using the large one. Given $C_1'(\mathbf{x},\mathbf{d})$ and $C_2'(\mathbf{x},\mathbf{d})$, the final filtered cost $C_f(\mathbf{x},\mathbf{d})$ at pixel $\mathbf{x}$ is set equal to the filtered cost that corresponds to the support window size that is appropriate for this pixel. The left

disparity map $d_{\mathrm{LR}}(\mathbf{x})$ (Figure 16a) is acquired after applying WTA to the cost volume $C_f(\mathbf{x},\mathbf{d})$. If the right image is considered as reference, then the disparity map $d_{\mathrm{RL}}(\mathbf{x})$ of Figure 16b is acquired.

In subsection 6.5.2.2, the selection of rectangular support windows of two sizes is experimentally justified. Provably, more than two window sizes could be used. However, this would increase the computational cost of the algorithm. Moreover, two window sizes are sufficient to achieve high disparity estimation accuracy.

4.3 Summary

This chapter described the matching cost computation and the cost aggregation steps for methodologies A and B.

In brief, methodology A used RGB information, weighted CENSUS transform and SIFT coefficients to define the pixel similarity measures $C_{\mathrm{RGB}}(\mathbf{x},\mathbf{d})$, $C_{\mathrm{CENSUS}}(\mathbf{x},\mathbf{d})$ and $C_{\mathrm{SIFT}}(\mathbf{x},\mathbf{d})$, respectively. These pixels measures were used to form a RGB-CENSUS combined matching cost $C_{\mathrm{R\text{-}C}}(\mathbf{x},\mathbf{d})$, a pure weighted CENSUS-based matching cost $C_{\mathrm{CENSUS}}(\mathbf{x},\mathbf{d})$ and a SIFT-based matching cost $C_{\mathrm{SIFT}}(\mathbf{x},\mathbf{d})$. After applying adaptive support-weight based aggregation for $C_{\mathrm{R\text{-}C}}(\mathbf{x},\mathbf{d})$, $C_{\mathrm{CEN}}(\mathbf{x},\mathbf{d})$ and $C_{\mathrm{S}}(\mathbf{x},\mathbf{d})$ the aggregated cost volumes $V_{\mathrm{R\text{-}C}}(\mathbf{x},\mathbf{d})$, $V_{\mathrm{CEN}}(\mathbf{x},\mathbf{d})$ and $V_{\mathrm{SIFT}}(\mathbf{x},\mathbf{d})$ were estimated, respectively. A novel two-phase strategy for combining $V_{\mathrm{R\text{-}C}}$, V_{CEN} and V_{SIFT} was introduced. During the first combination phase, $V_{\mathrm{CEN}}(\mathbf{x},\mathbf{d})$ was used to refine $V_{\mathrm{R\text{-}C}}(\mathbf{x},\mathbf{d})$ resulting to $V'_{\mathrm{R\text{-}C}}(\mathbf{x},\mathbf{d})$. While during the second combination phase, $V_{\mathrm{SIFT}}(\mathbf{x},\mathbf{d})$ was used to refine $V'_{\mathrm{R\text{-}C}}(\mathbf{x},\mathbf{d})$ resulting to $C_f(\mathbf{x},\mathbf{d})$ cost volume.

On the other hand, in methodology B, a gradient-based cost term $C_{\mathrm{gra}}(\mathbf{x},\mathbf{d})$, a Gabor-Feature-Image based term $C_{\mathrm{gab}}(\mathbf{x},\mathbf{d})$ and a Birchfield-Tomasi dissimilarity term $C_{\mathrm{BT}}(\mathbf{x},\mathbf{d})$ were linearly merged to form the combined matching cost $C(\mathbf{x},\mathbf{d})$. An innovative content-based guided image filtering approach was used to filter (aggregate) the matching costs $C(\mathbf{x},\mathbf{d})$. The content-based guided image filtering was applied separately for rectangular support windows of two different sizes and the

appropriate support window size for each pixel was selected based on the texture homogeneity within the local region around this pixel. After applying content-based guided image filtering to $C(\mathbf{x},\mathbf{d})$ the $C_f(\mathbf{x},\mathbf{d})$ cost volume was acquired.

In the forthcoming chapter, the disparity optimization and disparity refinement steps of methodology A and methodology B are extensively presented.

Chapter 5

5 Disparity optimization and disparity refinement

5.1 Disparity optimization

The disparity maps, which are acquired after performing the pixel-based matching cost and cost aggregation steps of Chapter 4, need further optimization in order to correct disparity estimation errors. Before applying disparity optimization, the outliers in problematic areas should be detected.

5.1.1 Outliers detection

The disparity maps $d_{\mathrm{LR}}(\mathbf{x})$ and $d_{\mathrm{RL}}(\mathbf{x})$ are taken into consideration to detect problematic areas, especially outliers in occluded regions and depth discontinuities. A prevalent strategy for detecting outliers is the Left-Right consistency check [44].

In this strategy, the outliers are disparity values that are not consistent between the two maps and therefore, they do not satisfy the relation:

$$\left|d_{\mathrm{LR}}(\mathbf{x}) - d_{\mathrm{RL}}(\mathbf{x} - d_{\mathrm{LR}}(\mathbf{x}))\right| \leq T_{\mathrm{LR}} \qquad (28)$$

5.1.1.1 Outliers detection for methodology A

In order to compute the outliers for methodology A, the disparity maps $d_{\mathrm{LR}}(\mathbf{x})$ (see Figure 14a) and $d_{\mathrm{RL}}(\mathbf{x})$ (see Figure 14b), which were computed in subsection 4.2.1.2, are considered in Equation (28).

The threshold for outliers detection is set equal to $T_{\mathrm{LR}} = 1$. With this value, pixels with difference equal to 1 in the Left-Right consistency check are not considered as outliers. This is plausible, since disparity in stereo images usually varies smoothly along the epipolar lines, in regions without depth discontinuities. Figure 17 shows the outliers map $\mathrm{O}_1^{T_{\mathrm{LR}}=1}(\mathbf{x})$ for $T_{\mathrm{LR}} = 1$. The blue regions are the outlier regions.

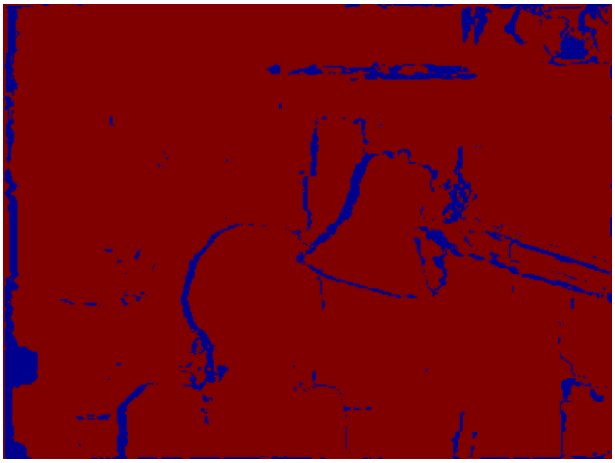


Figure 17. Outliers map $O_1^{T_{LR}=1}(\mathbf{x})$ in methodology A for threshold $T_{LR}=1$.

5.1.1.2 Outliers detection for methodology B

In order to compute the outliers for methodology B, the disparity maps $d_{LR}(\mathbf{x})$ (Figure 16a) and $d_{RL}(\mathbf{x})$ (Figure 16b), which were computed in subsection 4.2.2.2, are considered in Equation (28).

The threshold for outliers detection is set equal to $T_{LR}=0$. Figure 18 shows the outliers map $O_1^{T_{LR}=0}(\mathbf{x})$ for $T_{LR}=0$. The blue regions denote the outliers.

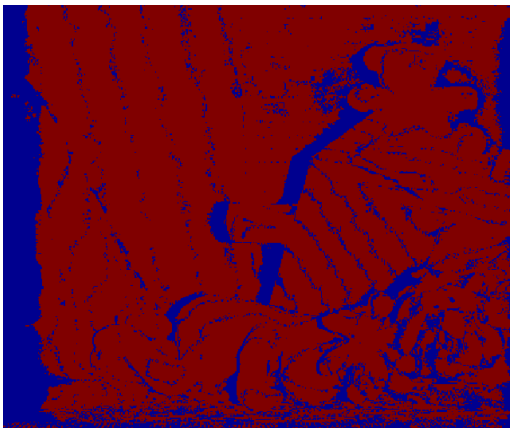


Figure 18. Outliers map $O_1^{T_{LR}=0}(\mathbf{x})$ in methodology B for threshold $T_{LR}=0$.

5.1.2 Enhanced semi-global disparity optimization

The disparity optimization relies on the semi-global optimization method of [38], which aggregates matching costs in 1D from multiple path directions. This subsection provides information on how to improve the accuracy of the original semi-

global optimization method.

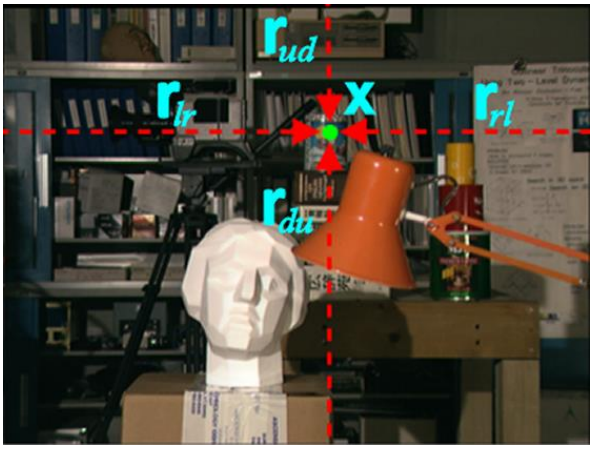


Figure 19. Path directions used for semi-global optimization.

The semi-global optimization approach considers four path directions $\mathbf{r}$, namely left-to-right, right-to-left, up-to-down and down-to-up, which are denoted as $\mathbf{r}_{lr}=[+1,0]^{\mathrm{T}}$, $\mathbf{r}_{rl}=[-1,0]^{\mathrm{T}}$, $\mathbf{r}_{ud}=[0,+1]^{\mathrm{T}}$ and $\mathbf{r}_{du}=[0,-1]^{\mathrm{T}}$, respectively (see Figure 19).

Let $L_{\mathbf{r}}$ be a path that is traversed in the direction $\mathbf{r}\in\{\mathbf{r}_{lr},\mathbf{r}_{rl},\mathbf{r}_{ud},\mathbf{r}_{du}\}$. The path cost $L_{\mathbf{r}}(\mathbf{x},\mathbf{d})$ of pixel $\mathbf{x}$ at disparity $\mathbf{d}$ is computed recursively from:

$$L_{\mathbf{r}}(\mathbf{x},\mathbf{d})=C_f(\mathbf{x},\mathbf{d})+\min\Big\{L_{\mathbf{r}}(\mathbf{x}-\mathbf{r},\mathbf{d}),L_{\mathbf{r}}(\mathbf{x}-\mathbf{r},\mathbf{d}\pm1)+\pi_1(\mathbf{x}),\\ \min_{\mathbf{d}_i}L_{\mathbf{r}}(\mathbf{x}-\mathbf{r},\mathbf{d}_i)+\pi_2(\mathbf{x})\Big\}-\min_{\mathbf{d}_i}L_{\mathbf{r}}(\mathbf{x}-\mathbf{r},\mathbf{d}_i) \quad (29)$$

where $\mathbf{d}_i\in[\textit{disparity range}]$ and $\mathbf{x}-\mathbf{r}$ denotes the previous pixel along the path direction. Parameters $\pi_1(\mathbf{x})$ and $\pi_2(\mathbf{x})$ are two smoothness penalty terms (with $\pi_1(\mathbf{x})\le\pi_2(\mathbf{x})$) for penalizing disparity changes between neighbouring pixels. The work in [44] assumes that a depth discontinuity usually coincides with an intensity edge; hence the smoothness penalty must be relaxed along edges and enforced within low-textured areas. Therefore, it applies a symmetrical strategy so that $\pi_1(\mathbf{x})$ and $\pi_2(\mathbf{x})$ depend on the intensities of both left and right images. In this PhD, two criteria are used to check depth discontinuity. The first criterion, similarly to [16], is based on intensity difference, which is computed as:

$$\nabla(\mathbf{x})=\left|I_l(\mathbf{x})-I_l(\mathbf{x}-\mathbf{r})\right| \quad (30)$$

and

$$\nabla(\mathbf{x}^{\mathbf{d}})=\left|I_r(\mathbf{x}^{\mathbf{d}})-I_r(\mathbf{x}^{\mathbf{d}}-\mathbf{r})\right|, \quad (31)$$

where I_l and I_r are the images in grayscale.

The second criterion, introduced in this PhD, checks whether two pixels belong to the same mean-shift segment. Let us assume that after applying mean-shift segmentation to the left and right images the label images Lab_l and Lab_r, are acquired. Each segment is denoted by a specific label. The second criterion is denoted as:

$$\Delta L_l(\mathbf{x}) = Lab_l(\mathbf{x}) - Lab_l(\mathbf{x}-\mathbf{r}) \quad (32)$$

and

$$\Delta L_r(\mathbf{x^d}) = Lab_r(\mathbf{x^d}) - Lab_r(\mathbf{x^d}-\mathbf{r}) \quad (33)$$

The smoothness penalty terms are defined according to:

$$\left(\pi_1(\mathbf{x}),\pi_2(\mathbf{x})\right) = \begin{cases} (\Pi_1,\Pi_2), \text{if } \left(\nabla(\mathbf{x}) \le \tau_{so} \ \& \ \nabla(\mathbf{x^d}) \le \tau_{so}\right) \\ \left(\frac{\Pi_1}{1.5},\frac{\Pi_2}{1.5}\right), \text{if } \left(\Delta L_l(\mathbf{x}) == 0 \ \& \ \Delta L_r(\mathbf{x^d}) == 0\right) \\ \left(\frac{\Pi_1}{4},\frac{\Pi_2}{4}\right), \text{if } \left(\nabla(\mathbf{x}) \le \tau_{so} \ \& \ \nabla(\mathbf{x^d}) > \tau_{so}\right) \ || \ \left(\Delta L_l(\mathbf{x}) == 0 \ \& \ \Delta L_r(\mathbf{x^d}) \ne 0\right) \\ \left(\frac{\Pi_1}{4},\frac{\Pi_2}{4}\right), \text{if } \left(\nabla(\mathbf{x}) > \tau_{so} \ \& \ \nabla(\mathbf{x^d}) \le \tau_{so}\right) \ || \ \left(\Delta L_l(\mathbf{x}) \ne 0 \ \& \ \Delta L_r(\mathbf{x^d}) == 0\right) \\ \left(\frac{\Pi_1}{10},\frac{\Pi_2}{10}\right), \text{otherwise} \end{cases} \quad (34)$$

where τ_{so} is a threshold for colour difference, Π_1, Π_2 are constant parameters and Lab_l, Lab_r are the labels images after applying mean-shift segmentation (see subsection 3.4.3) to the left and right images, respectively.

Existing methods, such as those in [16], [44] use only intensity based criteria to check intensity discontinuity and define parameters $\pi_1(\mathbf{x})$ and $\pi_2(\mathbf{x})$. The second criterion, which is based on mean-shift segmentation, improves the refinement results, as it is experimentally verified. The reason behind this improvement is that sometimes the first criterion denotes incorrectly a depth discontinuity due to texture edges that may be contained in image areas where depth does not change. On the contrary, mean-shift image segmentation is able to distinguish better between object texture edges and object boundaries. Therefore, the segmentation results are exploited for the definition of the smoothness penalties. In order to compensate for segmentation errors (include in the same segment areas with different depth) the denominator used for the definition of $\pi_1(\mathbf{x})$ and $\pi_2(\mathbf{x})$ is slightly increased to 1.5 for

the case that the second statement of Equation (34) is satisfied.

After computing the four path costs ($L_{\mathbf{r}_{lr}}, L_{\mathbf{r}_{rl}}, L_{\mathbf{r}_{ud}}, L_{\mathbf{r}_{du}}$) using Equation (29), the final cost volume $C_{\text{opt}}(\mathbf{x},\mathbf{d})$ is calculated from:

$$C_{\text{opt}}(\mathbf{x},\mathbf{d}) = \frac{w_{lr}(\mathbf{x})\cdot L_{\mathbf{r}_{lr}}(\mathbf{x},\mathbf{d}) + w_{rl}(\mathbf{x})\cdot L_{\mathbf{r}_{rl}}(\mathbf{x},\mathbf{d}) + w_{ud}(\mathbf{x})\cdot L_{\mathbf{r}_{ud}}(\mathbf{x},\mathbf{d}) + w_{du}(\mathbf{x})\cdot L_{\mathbf{r}_{du}}(\mathbf{x},\mathbf{d})}{4}, \quad (35)$$

where $w_{lr}(\mathbf{x}) + w_{rl}(\mathbf{x}) + w_{ud}(\mathbf{x}) + w_{du}(\mathbf{x}) = 4$.

In the original approach of the semi-global optimization [38]: $w_{lr}(\mathbf{x}) = w_{rl}(\mathbf{x}) = w_{ud}(\mathbf{x}) = w_{du}(\mathbf{x}) = 1$, while in the proposed modification these weights may not be equal. Practically, if along a path direction, the non-outlier pixels that belong to the same surface as the considered pixel $\mathbf{x}$, are much more than the non-outliers pixels of other directions, it is assumed that this direction should get a higher weight since it will give more accurate estimates. Therefore, for a pixel $\mathbf{x}$ and a specific direction, the total number of non-outlier pixels that precede $\mathbf{x}$ along this direction and at the same time they belong to the same surface as $\mathbf{x}$, is computed. The total number of non-outlier pixels for directions $\mathbf{r}_{lr}, \mathbf{r}_{rl}, \mathbf{r}_{ud}$ and $\mathbf{r}_{du}$ for pixel $\mathbf{x}$ is denoted as $\mathrm{M}_l'(\mathbf{x})$, $\mathrm{M}_r'(\mathbf{x})$, $\mathrm{M}_u'(\mathbf{x})$ and $\mathrm{M}_d'(\mathbf{x})$, respectively. $\mathrm{M}_l'(\mathbf{x})$, $\mathrm{M}_r'(\mathbf{x})$, $\mathrm{M}_u'(\mathbf{x})$ and $\mathrm{M}_d'(\mathbf{x})$ are computed as described in the next paragraph.

Let that the lengths of a pixel's arms, as estimated in subsection 4.2.2.2, are $\mathrm{M}_l(\mathbf{x})$, $\mathrm{M}_r(\mathbf{x})$, $\mathrm{M}_u(\mathbf{x})$ and $\mathrm{M}_d(\mathbf{x})$. The number of the pixels across an arm, which are outliers according to the outliers map $\mathrm{O}_1^{T_{\mathrm{LR}}=0}(\mathbf{x})$ (see Figure 18), is subtracted from the size of the arm. The sizes of the arms, after subtracting the number of outlier pixels, are denoted as $\mathrm{M}_l'(\mathbf{x})$, $\mathrm{M}_r'(\mathbf{x})$, $\mathrm{M}_u'(\mathbf{x})$ and $\mathrm{M}_d'(\mathbf{x})$, respectively.

Let $\mathrm{M}_{\max}'(\mathbf{x})$ denote the maximum value among $\mathrm{M}_l'(\mathbf{x})$, $\mathrm{M}_r'(\mathbf{x})$, $\mathrm{M}_u'(\mathbf{x})$ and $\mathrm{M}_d'(\mathbf{x})$, while $\mathrm{M}_{\mathrm{sec}}'(\mathbf{x})$ denotes the second highest value. Based on $\mathrm{M}_{\max}'(\mathbf{x})$ and $\mathrm{M}_{\mathrm{sec}}'(\mathbf{x})$, the following conditions are defined:

$$\mathrm{M}_{\max}'(\mathbf{x})/\mathrm{M}_{\mathrm{sec}}'(\mathbf{x}) > 2 \ \text{ and } \ \mathrm{M}_{\max}'(\mathbf{x}) > \mathrm{R_S}/2 \quad (36)$$

The first condition confirms that a direction has much more non-outlier pixels than the other directions, while the second condition confirms that there is a sufficient number of non-outlier pixels along this direction. In case both the conditions

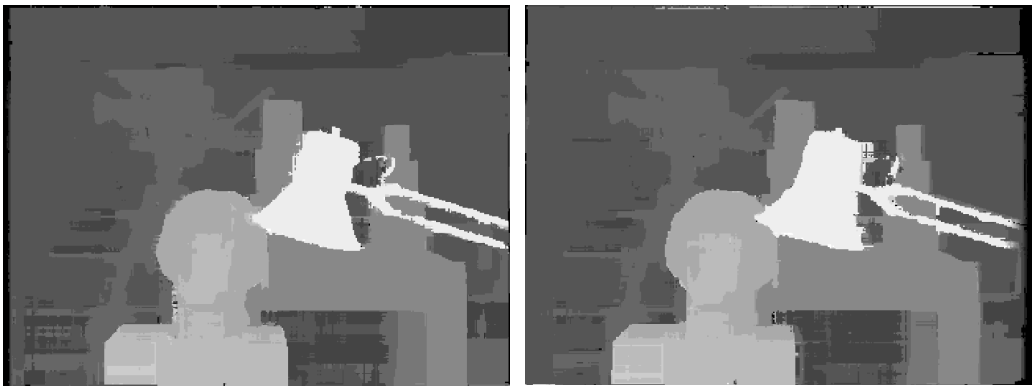
in Equation (36) are satisfied, then a higher weight is given to the path cost that corresponds to the direction from which $\mathrm{M}_{\max}^{'}(\mathbf{x})$ has been derived.

For example, if $\mathrm{M}_{\max}^{'}(\mathbf{x})$ is equal to $\mathrm{M}_{u}^{'}(\mathbf{x})$, which corresponds to direction $\mathbf{r}_{ud}$, then the weights used in Equation (35) will be set as: $w_{ud}(\mathbf{x})=1.6$ and $w_{lr}(\mathbf{x})=w_{rl}(\mathbf{x})=w_{du}(\mathbf{x})=0.8$. That is, a higher weight is given to the direction that has much more pixels that belong to the same surface as $\mathbf{x}$, when compared to the other directions, which at the same time are non-outliers. If any of the conditions in Equation (36) is not satisfied then all weights are set equal to 1.

To summarize, two novel ideas, regarding the semi-global optimization, have been introduced in this subsection. The first idea, which is used by methodology A, concerns the introduction of a criterion that relies on mean-shift segmentation for the detection of depth discontinuities. The second idea, which is used by methodology B, concerns the introduction of a scheme for defining the weights of each path cost.

5.1.2.1 Disparity maps after optimization for methodology A

Having as input in Equation (29) the $C_f(\mathbf{x},\mathbf{d})$ volume, which has been estimated in subsection 4.2.1.2, the optimized cost volume $C_{\text{opt}}(\mathbf{x},\mathbf{d})$ is acquired via equation (35). The WTA of $C_{\text{opt}}(\mathbf{x},\mathbf{d})$ gives the disparity map $d_{\text{LR}}^{'}(\mathbf{x})$ (see Figure 20a). If the right image is considered as reference image, then the disparity map $d_{\text{RL}}^{'}(\mathbf{x})$ (see Figure 20b) is acquired.

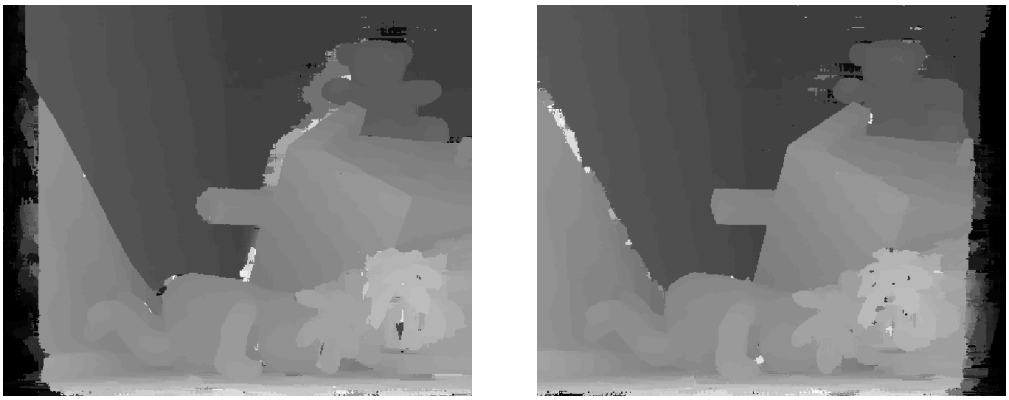


(a) (b)

Figure 20. Disparity maps (a) $d_{\text{LR}}^{'}(\mathbf{x})$ and (b) $d_{\text{RL}}^{'}(\mathbf{x})$ after optimization for methodology A.

5.1.2.2 Disparity maps after optimization for methodology B

Having as input in Equation (29) the $C_f(\mathbf{x},\mathbf{d})$ volume, which has been estimated in subsection 4.2.2.2, the optimized cost volume $C_{\text{opt}}(\mathbf{x},\mathbf{d})$ is acquired via equation (35). The WTA of $C_{\text{opt}}(\mathbf{x},\mathbf{d})$ gives the disparity map $d'_{\text{LR}}(\mathbf{x})$ (see Figure 21a). If the right image is considered as reference image, then the disparity map $d'_{\text{RL}}(\mathbf{x})$ (see Figure 21b) is acquired.



(a) **(b)**

Figure 21. Disparity maps (a) $d'_{\text{LR}}(\mathbf{x})$ and (b) $d'_{\text{RL}}(\mathbf{x})$ after optimization for methodology B.

5.2 Disparity refinement

The optimized disparity results have to be refined, since they are polluted with outliers in occluded areas, low-textured areas and depth discontinuities. This section provides detail on how the outliers can be handled after they are detected.

5.2.1 Outliers detection from optimized disparity maps

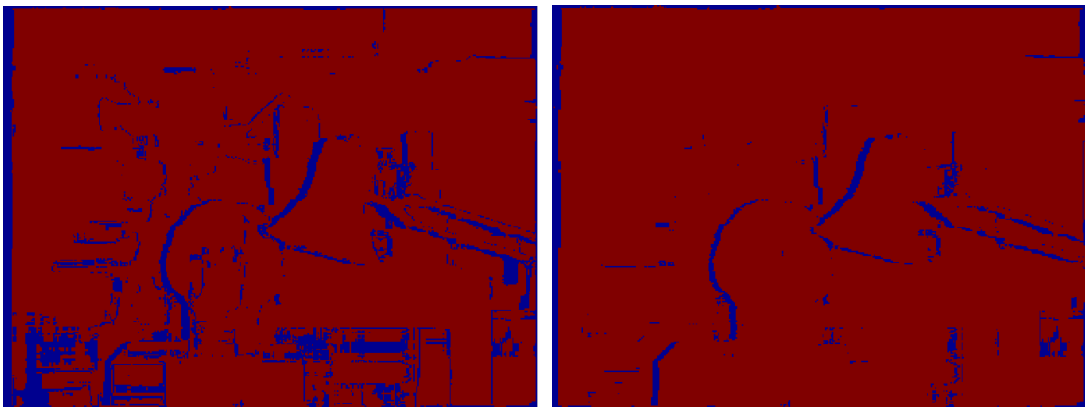
The optimized disparity maps $d'_{\text{LR}}(\mathbf{x})$ and $d'_{\text{RL}}(\mathbf{x})$ are taken into consideration to detect problematic areas. The outlier pixels do not satisfy the relation:

$$\left|d'_{\text{LR}}(\mathbf{x}) - d'_{\text{RL}}(\mathbf{x} - d'_{\text{LR}}(\mathbf{x}))\right| \le T_{\text{LR}} \quad (37)$$

5.2.1.1 Outliers detection for methodology A

In order to compute the outliers for methodology A, the disparity maps $d'_{\text{LR}}(\mathbf{x})$ (see Figure 20a) and $d'_{\text{RL}}(\mathbf{x})$ (see Figure 20a), are considered in Equation (37).

For $T_{\mathrm{LR}}=0$ and $T_{\mathrm{LR}}=1$ the outliers maps $\mathrm{O}_2^{T_{\mathrm{LR}}=0}(\mathbf{x})$ (see Figure 22a) and $\mathrm{O}_2^{T_{\mathrm{LR}}=1}(\mathbf{x})$ (see Figure 22b) are acquired, respectively.



(a) (b)

Figure 22. Outliers map (a) $\mathrm{O}_2^{T_{\mathrm{LR}}=0}(\mathbf{x})$ for threshold $T_{\mathrm{LR}}=0$ and (b) $\mathrm{O}_2^{T_{\mathrm{LR}}=1}(\mathbf{x})$ for threshold $T_{\mathrm{LR}}=1$.

5.2.1.2 Outliers detection for methodology B

In order to compute the outliers for methodology B, the disparity maps $d_{\mathrm{LR}}^{'}(\mathbf{x})$ (see Figure 21a) and $d_{\mathrm{RL}}^{'}(\mathbf{x})$ (Figure 21b), are considered in Equation (37). For $T_{\mathrm{LR}}=0$ the outlier map $\mathrm{O}_2^{T_{\mathrm{LR}}}(\mathbf{x})$ (see Figure 23) is acquired.

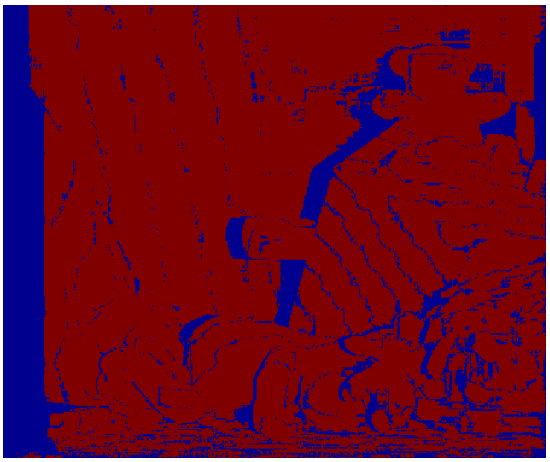


Figure 23. Outliers map $\mathrm{O}_2^{T_{\mathrm{LR}}}(\mathbf{x})$ for threshold $T_{\mathrm{LR}}=0$.

5.2.2 Outliers handling

With the algorithmic steps, described through this subsection, the outliers

that are contained in occluded regions, uniform areas and depth discontinuities can be efficiently handled.

5.2.2.1 Outliers handling for methodology A

Outliers handling in methodology A is performed by combining two outlier handling schemes that are executed independently. The first outlier scheme is called "Basic outlier handling'' and the second scheme is called "Mean-shift segmentation-based outlier handling".

Basic outlier handling

The basic outlier handling strategy is performed for the outlier map $\mathrm{O}_2^{T_{\mathrm{LR}}=0}(\mathbf{x})$ (see Figure 22a). In more detail, an outlier pixel $\mathbf{x}$ is filled by the disparity of its closest inlier pixel. Practically, the disparity values of $\mathbf{x}$'s left nearest inlier pixel $\mathbf{x}_l$ and $\mathbf{x}$'s right nearest inlier pixel $\mathbf{x}_r$ are denoted as $d_{\mathrm{LR}}^{'}(\mathbf{x}_l)$ and $d_{\mathrm{LR}}^{'}(\mathbf{x}_r)$, respectively. Then, the disparity value of $\min(d_{\mathrm{LR}}^{'}(\mathbf{x}_l), d_{\mathrm{LR}}^{'}(\mathbf{x}_r))$ is assigned to $\mathbf{x}$. The disparity map, after the basic outlier handling, is visualized in Figure 24c.

Mean-shift segmentation-based outlier handling

There is high probability that the candidate ``outlier'' points for $T_{\mathrm{LR}}=1$, are not actual outliers. Instead, it is probable that there is a slight difference in the disparity estimation between the left and the right disparity maps. The following technique is applied to propagate reliably disparity information from the right disparity map to the left disparity map.

For a pixel ϕ, with $T_{\mathrm{LR}}(\phi)=1$, the subset of pixels within radius 7 from ϕ, which at the same time belong to the same segment as ϕ, is defined. This subset is used to estimate a reliability metric $Rel_{\mathrm{LR}}(\phi)$, whose value is given by the division of the number of pixels within this subset with $T_{\mathrm{LR}}=0$ towards the total number of pixels in this subset. Correspondingly, for a pixel ψ, which is the correspondence of pixel ψ in the right image, the metric $Rel_{\mathrm{RL}}(\psi)$ is similarly computed.

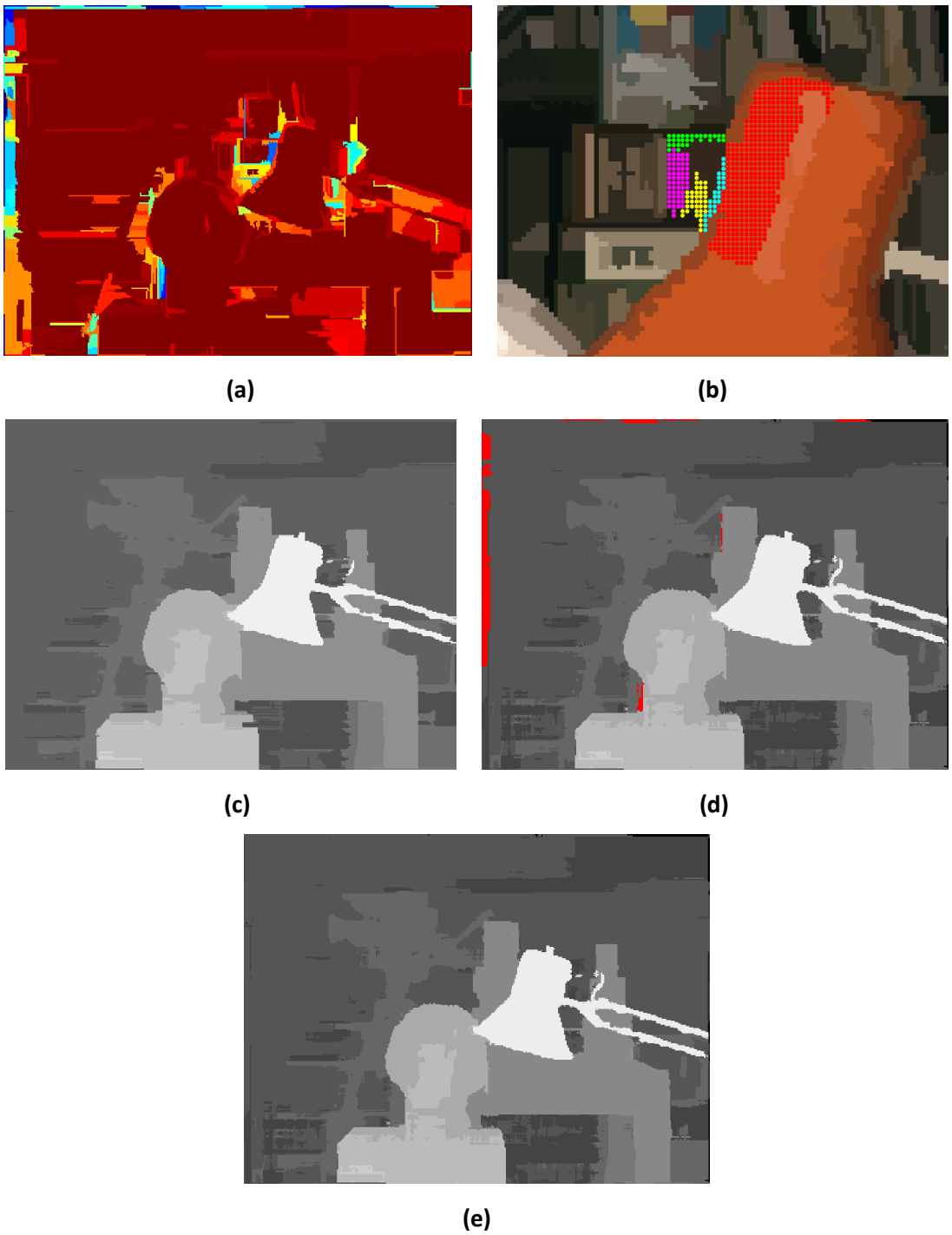


Figure 24. Illustration of: (a) the reliability map, (b) an unreliable segment and its neighboring segments, (c) the disparity map after applying basic outlier handling, (d) the disparity map mean-shift based segmentation outlier handling, (e) the disparity map after combined outlier handling.

The disparity of pixel ψ is propagated to pixel ϕ in case that $d'_{\mathrm{LR}}(\phi) > d'_{\mathrm{RL}}(\psi)$ and $Rel_{\mathrm{LR}}(\phi) < Rel_{\mathrm{RL}}(\psi)$. Pixels ϕ, whose disparity has been propagated from their corresponding ψ pixels and the pixels ϕ with $T_{\mathrm{LR}}(\phi) = 0$ are considered as

"unoccluded'". These "unoccluded" pixels are used in the application of the mean shift segmentation-based outlier handling, as follows.

Initially, for each mean-shift segment the ratio of the unoccluded pixels inside this segment over the total number of segment's pixels is evaluated. This ratio constitutes a reliability measure for the disparities inside this segment. Such a reliability map is illustrated in Figure 24a. The warmer the colour, the more reliable the disparities inside a segment are. A segment is considered as "reliable" if the ratio is over T_{r} (experimentally defined to be 0.3).

Reliable segments: For the outlier pixels inside a reliable segment S, a voting scheme that counts votes of the reliable pixels' disparities is applied. In more detail, for each outlier pixel $\mathbf{x} \in S$, the inlier pixels that belong to S and lie within Euclidean distance R_{S} (radius of support region defined in subsection 4.2.1.1) from $\mathbf{x}$ are taken into account in order to get the most frequent disparity. This disparity is propagated to $\mathbf{x}$ which is considered as reliable now. This process is repeated for all outliers inside a segment S.

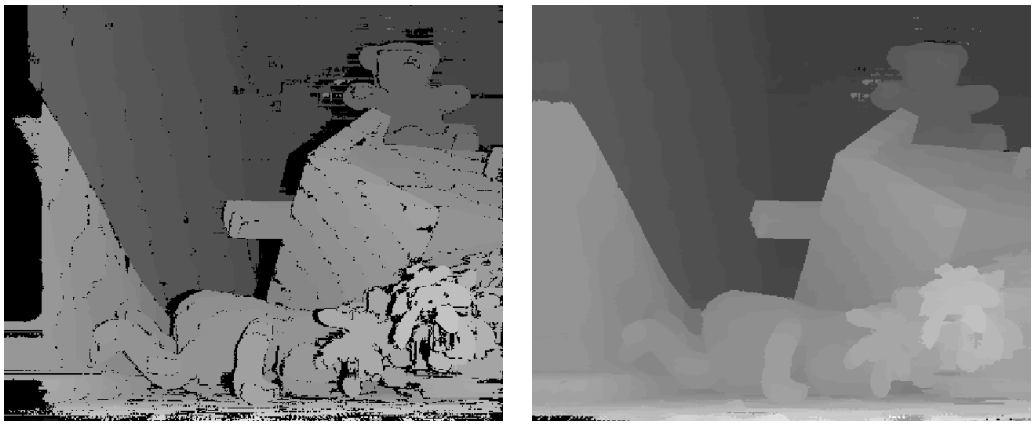
Unreliable segments: For unreliable segments, the information from reliable neighboring segments is used to define their disparity. Reliable neighboring segments are the reliable segments that have common borders with the unreliable segments. For example, in Figure 24b the unreliable segment is surrounded by the colored neighboring segments. The reliable neighboring segment that will propagate its prevalent disparity to the unreliable segment is the one that has the most similar colour to the unreliable segment. Notice that the mean colour of each segment was estimated during the mean-shift segmentation. The colour similarity between two segments is defined as the mean Euclidean distance between their mean RGB colours and should be below T_{S} (experimentally defined to be 25). The disparity map, after handling outlier areas based on the mean-shift segmentation-based outlier handling scheme, is visualized in Figure 24d. The red areas in Figure 24d correspond to pixels that have not been handled using the mean-shift segmentation-based outlier handling.

Combined outlier handling

Finally, the occluded areas that have not been handled using the mean-shift based segmentation outlier handling are filled with the disparities that have been estimated through the basic outlier handling and in this way the combined disparity map of Figure 24e is acquired.

5.2.2.2 Outliers handling for methodology B

Outliers handling in methodology B is performed by combining "Background outliers handling" and "Generic outliers handling".



(a) **(b)**

Figure 25. Disparity map after applying: (a) "Generic outliers handling" and (b) "Background outliers handling" plus bilateral smoothing.

Background outliers handling

The ``Background outliers handling'' is performed for the outlier map $\mathrm{O}_2^{T_{\mathrm{LR}}}(\mathbf{x})$ (see Figure 23). One of the simplest approaches for handling an outlier pixel $\mathbf{x}$, which belongs to the pixels of the occluded background, is to set its disparity $d_{\mathrm{LR}}^{'}(\mathbf{x})$ equal to the disparity of its closest consistent pixel [18]. That is, the minimum between $d_{\mathrm{LR}}^{'}(\mathbf{x}_l)$ and $d_{\mathrm{LR}}^{'}(\mathbf{x}_r)$ is assigned to $d_{\mathrm{LR}}^{'}(\mathbf{x})$, where $\mathbf{x}_l$ and $\mathbf{x}_r$ stand for the nearest consistent pixels on the left and the right of pixel $\mathbf{x}$, respectively.

Generic outliers handling

Since the outliers do not always correspond to background occlusions, it has been introduced a straightforward scheme which precedes the ``Background outliers handling''. This scheme does not presume that an outlier pixel $\mathbf{x}$ belongs to the background, but it checks whether its left or right side has more similar (in term of colour) pixels to that pixel. In more detail, for an outlier pixel $\mathbf{x}$, separately for the left and right side, the inlier pixels, for which the following condition is verified, are counted:

$$\left|I_l(\mathbf{x}) - I_l(\mathbf{x}+\mathbf{s})\right| < \tau_1, \tag{38}$$

where $\mathbf{s} = (-\mathrm{s}_{\mathrm{x}_l}, 0)^{\mathbf{T}}$, $\mathrm{s}_{\mathrm{x}_l} \in [1, \mathrm{s}_{l_{\max}}(\mathbf{x}))$ for the left side and $\mathbf{s} = (\mathrm{s}_{\mathrm{x}_r}, 0)^{\mathbf{T}}$, $\mathrm{s}_{\mathrm{x}_r} \in [1, \mathrm{s}_{r_{\max}}(\mathbf{x}))$ for the right side, while τ_1 is a threshold for colour difference. $\mathrm{s}_{l_{\max}}(\mathbf{x})$ and $\mathrm{s}_{r_{\max}}(\mathbf{x})$ are the integer values for which the condition of Equation (38) fails for the first time when examining the left and the right sides, respectively. For the pixels on the left side of $\mathbf{x}$, the weights $\beta_l(\mathbf{x}+\mathbf{s})$ are calculated from:

$$\beta_l(\mathbf{x}+\mathbf{s}) = \begin{cases} 1, \text{if } \mathbf{x}+\mathbf{s} \text{ is inlier} \\ 0, \text{if } \mathbf{x}+\mathbf{s} \text{ is outlier.} \end{cases} \tag{39}$$

Afterwards, for the left side the following disparity histogram is generated:

$$\mathrm{H}_l(\mathbf{x}, d_{\mathrm{i}}) = \sum_{\mathbf{x}+\mathbf{s}:\, d(\mathbf{x}+\mathbf{s}) == d_{\mathrm{i}}} \beta_l(\mathbf{x}+\mathbf{s}), \tag{40}$$

where $d_{\mathrm{i}} \in [\textit{disparity range}]$.

In an analogous manner, the disparity histogram $\mathrm{H}_r(\mathbf{x}, d_{\mathrm{i}})$ for the right side is generated. Let now the maximum values of the left and the right histograms be $\mathrm{h}_{l_{\max}}(\mathbf{x}) = \max_{d_{\mathrm{i}}}\{\mathrm{H}_l(\mathbf{x}, d_{\mathrm{i}})\}$ and $\mathrm{h}_{r_{\max}}(\mathbf{x})$, respectively and the corresponding disparity values be $d_{l_{\max}}(\mathbf{x}) = \arg\max_{d_{\mathrm{i}}}\{\mathrm{H}_l(\mathbf{x}, d_{\mathrm{i}})\}$ and $d_{r_{\max}}(\mathbf{x})$, respectively. Based on the above, the new disparity estimate $d(\mathbf{x})$ is given from:

$$d(\mathbf{x}) = \begin{cases} d_{l_{\max}}(\mathbf{x}), \text{if } \left(\mathrm{h}_{l_{\max}}(\mathbf{x}) > \mathrm{h}_{r_{\max}}(\mathbf{x}) \,\&\, \mathrm{h}_{l_{\max}}(\mathbf{x}) > \mathrm{R}_{\mathrm{S}}/2\right) \\ d_{r_{\max}}(\mathbf{x}), \text{if } \left(\mathrm{h}_{l_{\max}}(\mathbf{x}) < \mathrm{h}_{r_{\max}}(\mathbf{x}) \,\&\, \mathrm{h}_{r_{\max}}(\mathbf{x}) > \mathrm{R}_{\mathrm{S}}/2\right) \end{cases} \tag{41}$$

The outliers that have been handled using Equation (41) are considered now

as inliers. The disparity map of Figure 21a after applying the ``Generic outliers handling'' is visualized in Figure 25a. For the remaining outliers (i.e. none of the conditions in Equation (41) holds), the approach in paragraph ``Background outliers handling'' is performed.

In order to deal with horizontal artifacts that are produced after applying the ``Background outliers handling'', a bilateral filter is used to smooth the filled regions. The bilateral filter weights are given by [18]:

$$W_{\mathbf{x},\mathbf{q}} = \frac{1}{k} \cdot \exp\left(-\left(\frac{\Delta s_{\mathbf{x},\mathbf{q}}}{\gamma_s} + \frac{\Delta c_{\mathbf{x},\mathbf{q}}}{\gamma_c} \right) \right), \quad (42)$$

where k is a normalization factor, $\Delta s_{\mathbf{x},\mathbf{q}}$ and $\Delta c_{\mathbf{x},\mathbf{q}}$ denote the spatial distance and the colour difference between pixels $\mathbf{x}$, $\mathbf{q}$ and γ_s, γ_c are constant parameters that adjust the spatial and colour distance. The parameters of the bilateral filter are set as in [18]: $\gamma_s = 9$, $\gamma_c = 0.1$ and its window size is $R_S \times R_S$. The disparity map of Figure 25a after applying ``Background outliers handling'' and bilateral smoothing is visualized in Figure 25b.

5.2.3 Disparity edges refinement

Disparity edges, which correspond to depth discontinuities, may contain disparity errors [16]. In the following, two simple approaches, which are used by the two methodologies to refine depth discontinuities, are presented.

5.2.3.1 Disparity edges refinement in methodology A

A two-step approach is used to refine the disparity information at the disparity edges. The first step detects and handles the disparity edges at a coarser level and the second one at a finer level.

The pixels that belong to a disparity edge are assumed to have a difference greater or equal to 2 with at least one of their 4-adjacent pixels disparity. Otherwise, if the difference is below 2, then the surface varies smoothly and therefore one can assume that there is no depth discontinuity. Figure 26b shows the disparity edges extracted from the disparity map of Figure 26a.

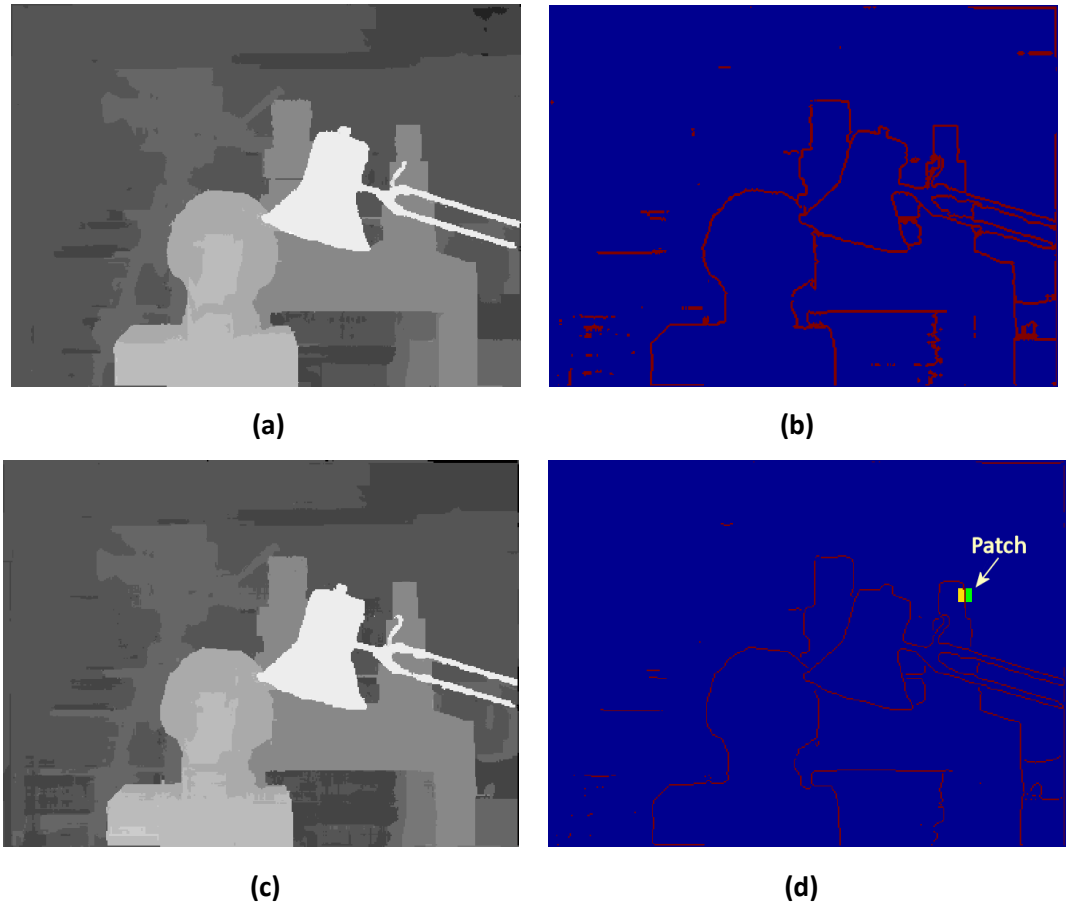


Figure 26. Illustration of: (a) the disparity map to be used for disparity edges refinements, (b) the disparity map's disparity edges, (c) the disparity map after coarse discontinuity refinement, (d) the disparity edges after applying canny disparity edge detection.

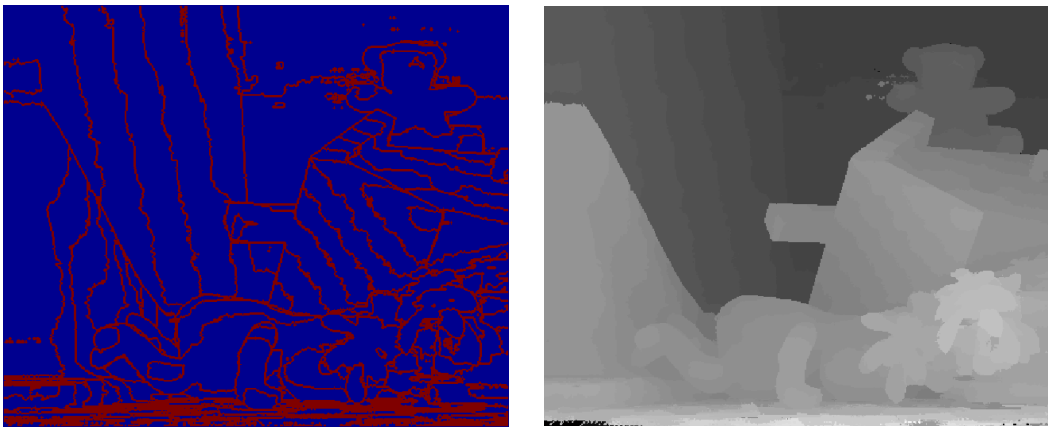
During the first step, around each pixel of the disparity edge, a circular region of radius 3 is defined. The disparities of the pixels that fall inside the circular region and at the same time belong to the same mean-shift segment, as the pixel of the disparity edge, are used to find the most frequent disparity value. This value is propagated to the edge pixel. The disparity result after the first step is depicted in Figure 26c.

The second step handles discontinuities at a finer scale. Firstly, canny edge detection (see Figure 26d) is applied to the disparity result of Figure 26c. Canny can detect disparity edges at finer scale than the coarse previously-described step (this is evident when comparing Figure 26b and Figure 26d). Then a patch of size 3x3 is centered at each edge point and the disparity regions separated by the edge are found. Figure 26d shows that the edge separates the patch into a yellow and green disparity

region. The disparity region that contains the pixel with the greatest colour similarity to the edge pixel (the colour similarity is found according to the initial reference stereo image) gives its disparity to the considered pixel.

5.2.3.2 Disparity edges refinement in methodology B

Methodology B uses a straightforward efficient approach to refine the disparity estimation at the edges. Initially, the pixels that belong to disparity edges are assumed to have absolute disparity difference greater or equal to 1 with at least one of their 4-adjacent pixels. Figure 27a shows the disparity edges extracted from the disparity map of Figure 25b.



(a) **(b)**

Figure 27. Illustration of: (a) the disparity edges of the disparity map to be used for disparity edges refinements, (b) the disparity map after disparity edges refinement.

Around each pixel $\mathbf{x}_c$ of the disparity edge, a circular region of radius 4 is defined. The colour similarity between the center pixel $\mathbf{x}_c$ and a pixel $\mathbf{x}$ within the circular region is estimated as:

$$w(\mathbf{x}_c, \mathbf{q}) = e^{\left(\frac{-\Delta I(\mathbf{x}_c, \mathbf{q})}{\gamma_c}\right)}, \quad (43)$$

where

$$\Delta I(\mathbf{x}_c, \mathbf{q}) = \sqrt{\sum_{c\in\{R,G,B\}} |I^c(\mathbf{x}_c) - I^c(\mathbf{q})|^2}. \quad (44)$$

A disparity histogram is generated for each $\mathbf{x}_c$, where the values of its disparity bins are computed as follows:

$$\mathrm{H}_{\mathbf{x}_\mathrm{c}}(\mathbf{x}_\mathrm{c}, d_\mathrm{i}) = \sum_{\mathbf{q}: d(\mathbf{q}) == d_\mathrm{i}} w(\mathbf{x}_\mathrm{c}, \mathbf{q}), \quad (45)$$

where $d_\mathrm{i} \in [\textit{disparity range}]$. Let now the maximum and the second maximum value of $\mathrm{H}_{\mathbf{x}_\mathrm{c}}(\mathbf{x}_\mathrm{c}, d_\mathrm{i})$ be $\mathrm{h}_{\max}(\mathbf{x}_\mathrm{c}) = \max_{d_\mathrm{i}}\{\mathrm{H}_{\mathbf{x}_\mathrm{c}}(\mathbf{x}_\mathrm{c}, d_\mathrm{i})\}$ and $\mathrm{h}_{\mathrm{sec}}(\mathbf{x}_\mathrm{c})$, respectively and the corresponding disparity value for $\mathrm{h}_{\max}(\mathbf{x}_\mathrm{c})$ be $d_{\mathrm{h}_{\max}}(\mathbf{x}_\mathrm{c}) = \arg\max_{d_\mathrm{i}}\{\mathrm{H}_{\mathbf{x}_\mathrm{c}}(\mathbf{x}_\mathrm{c}, d_\mathrm{i})\}$. If $\mathrm{h}_{\max}(\mathbf{x}_\mathrm{c}) / \mathrm{h}_{\mathrm{sec}}(\mathbf{x}_\mathrm{c}) > 2$ then $d(\mathbf{x}_\mathrm{c}) = d_{\mathrm{h}_{\max}}(\mathbf{x}_\mathrm{c})$, otherwise the disparity value of $d(\mathbf{x}_\mathrm{c})$ does not change.

The disparity result after the disparity edges refinement is depicted in Figure 27b. A median filter, using a 3x3 neighborhood, is applied to the disparity result of Figure 27b in order to remove spurious disparities before acquiring the final disparity map, which is depicted in the upper image of the third column Figure 35.

5.2.4 Selective uniform areas handling used in methodology A

Usually, images contain large uniform areas, where it is difficult to establish accurate pixel correspondences between two images. In order to deal with ambiguous matches in these areas, methodology A uses a novel approach presented in this subsection.

5.2.4.1 Detection of uniform areas

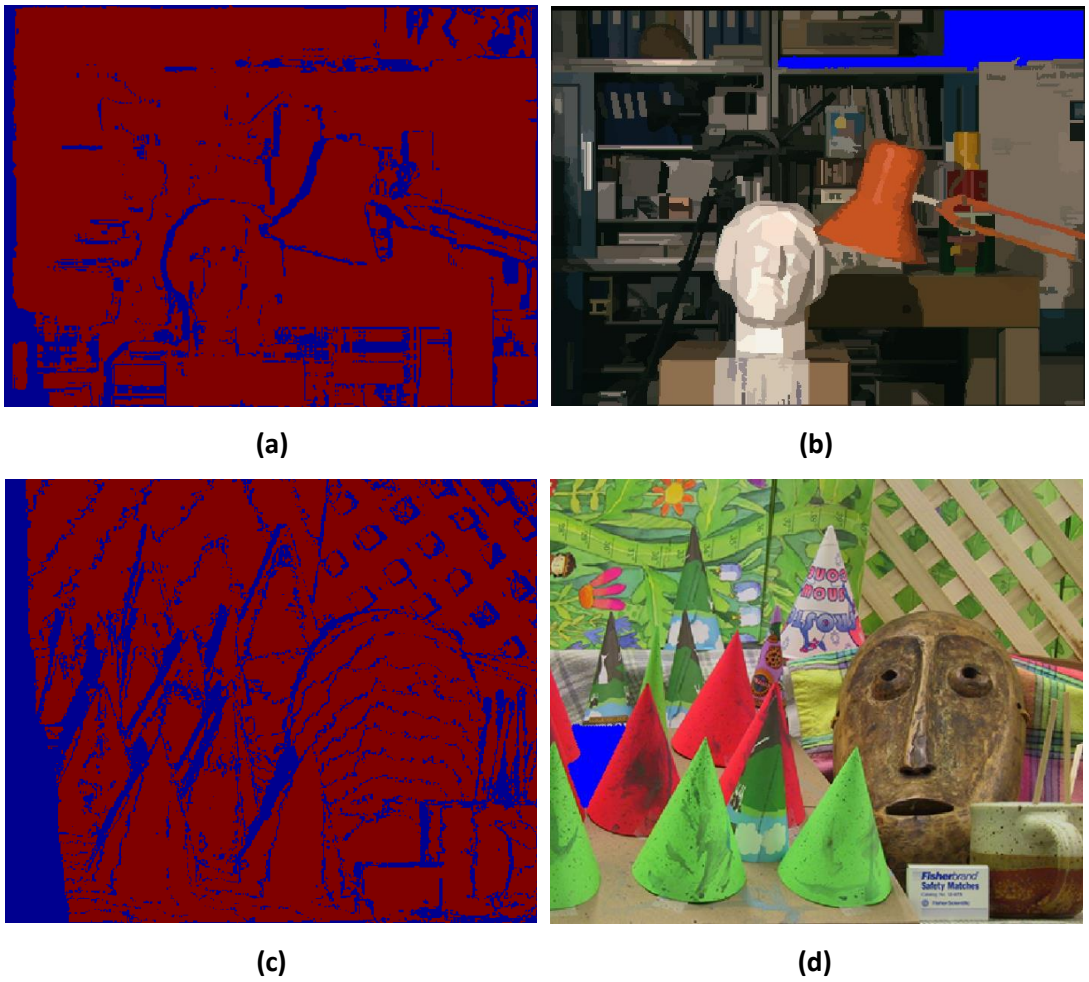
Initially, large uniform areas on the image are detected. Large uniform areas are considered to be the mean-shift segments that contain over $2 \cdot R_\mathrm{S}^2$ pixels (R_S is the radius of the support region as defined in subsection 4.2.1.1). Then, each segment's ``inlier'' pixels are estimated and used for the uniform areas handling.

5.2.4.2 Inlier pixels regions

Inlier pixels $\mathbf{x}$ from $d_{\mathrm{LR}}(\mathbf{x})$ (see Figure 14a in subsection 4.2.1.2), are used for the uniform-areas handling. The inlier pixels regions are determined as follows:

- The outliers map $\mathrm{O}_1^{T_{\mathrm{LR}}=1}(\mathbf{x})$ of subsection 5.1.1.1 (see Figure 17), as well as the

outliers map $O_2^{T_{LR}=0}(\mathbf{x})$ of subsection 5.2.1.1 (see Figure 22a) are considered in order to acquire their union, which defines the overall outliers map $O_U(\mathbf{x})$. In $O_U(\mathbf{x})$, outlier pixels are those that are outliers in either $O_1^{T_{LR}=1}(\mathbf{x})$ or $O_2^{T_{LR}=0}(\mathbf{x})$. Let X_{In} be the set of inlier pixels in $O_U(\mathbf{x})$.

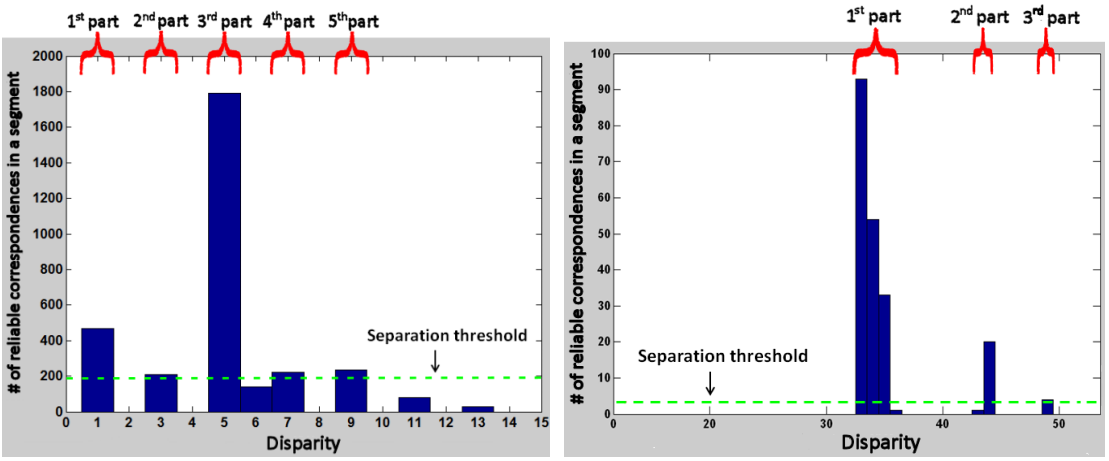


(a) **(b)**

(c) **(d)**

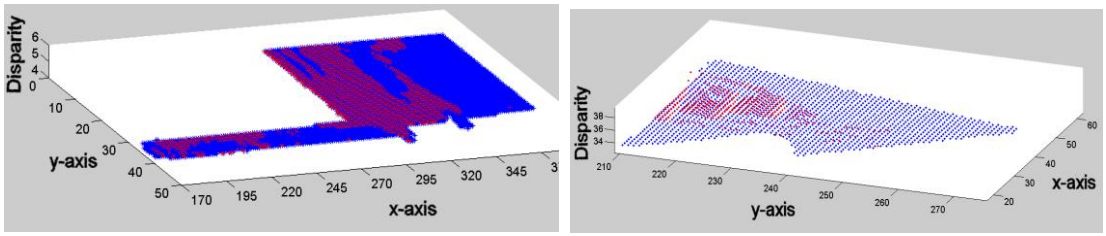
Figure 28. Inlier pixels (red regions) in $O_U(\mathbf{x})$ for (a) the left "Tsukuba" image and (c) the left "Cones" image. A segment on (b) the left "Tsukuba" image and (d) the left "Cones" image (shown with blue).

A visual example is given in the first row of Figure 28. Figure 28a shows the overall outliers map $O_U(\mathbf{x})$ that is generated after the union of the outliers maps $O_1^{T_{LR}=1}(\mathbf{x})$ and $O_2^{T_{LR}=0}(\mathbf{x})$ acquired in subsections 5.1.1.1 and 5.2.1.1, respectively. The inlier pixels X_{In} are denoted with red colour.



(a) (b)

Figure 29. Disparity histogram of the inlier pixels in a segment on (a) the left "Tsukuba" image and (b) the left "Cones" image.



(a) (b)

Figure 30. Fit a plane (blue) to a segment, applying PCA on the reliable subset of pixels (red) for a segment on (a) the left "Tsukuba" image and (b) the left "Cones" image.

5.2.4.3 Extraction of a reliable pixels, based on histogram analysis

A histogram analysis, based on the inlier pixels' disparities $d_{\mathrm{LR}}(\mathrm{X}_{\mathrm{In}})$ is applied in order to acquire a reliable subset of the pixels. For instance, for the mean-shift segment of Figure 28b (marked with blue colour), the histogram of the disparities of the inlier pixels inside this segment is depicted in Figure 29a.

Theoretically, the disparities of the pixels in a segment S should vary continuously within a disparity range, since they belong to the same continuous surface. Based on this assumption, the employed approach is followed to get the subset of the reliable pixels.

Initially, the histogram of disparities is separated into parts (each part expresses a disparity range), as shown in Figure 29a. To separate the histogram into

parts, bins with a height below a ``separation threshold'' are ignored, so that they do not affect the separation process. This threshold is selected equal to: $\frac{\text{Number of inlier pixels in } S}{\text{Number of possible disparities}}$. The reliable subset of inlier pixels includes the pixels whose disparities belong to the histogram part with the most numerous population (3rd part of Figure 29a).

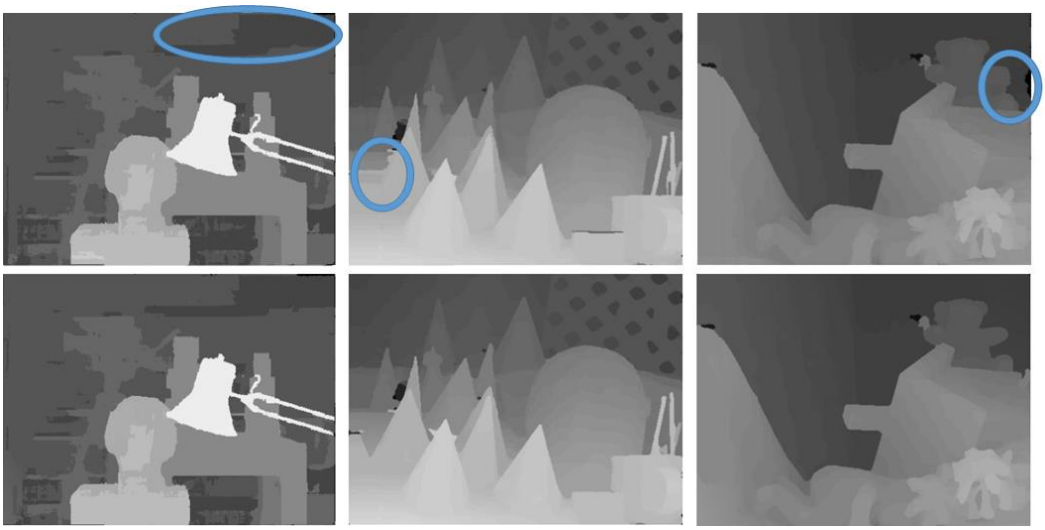


Figure 31. Disparity maps before (1st row) and after applying uniform region handling (2nd row).

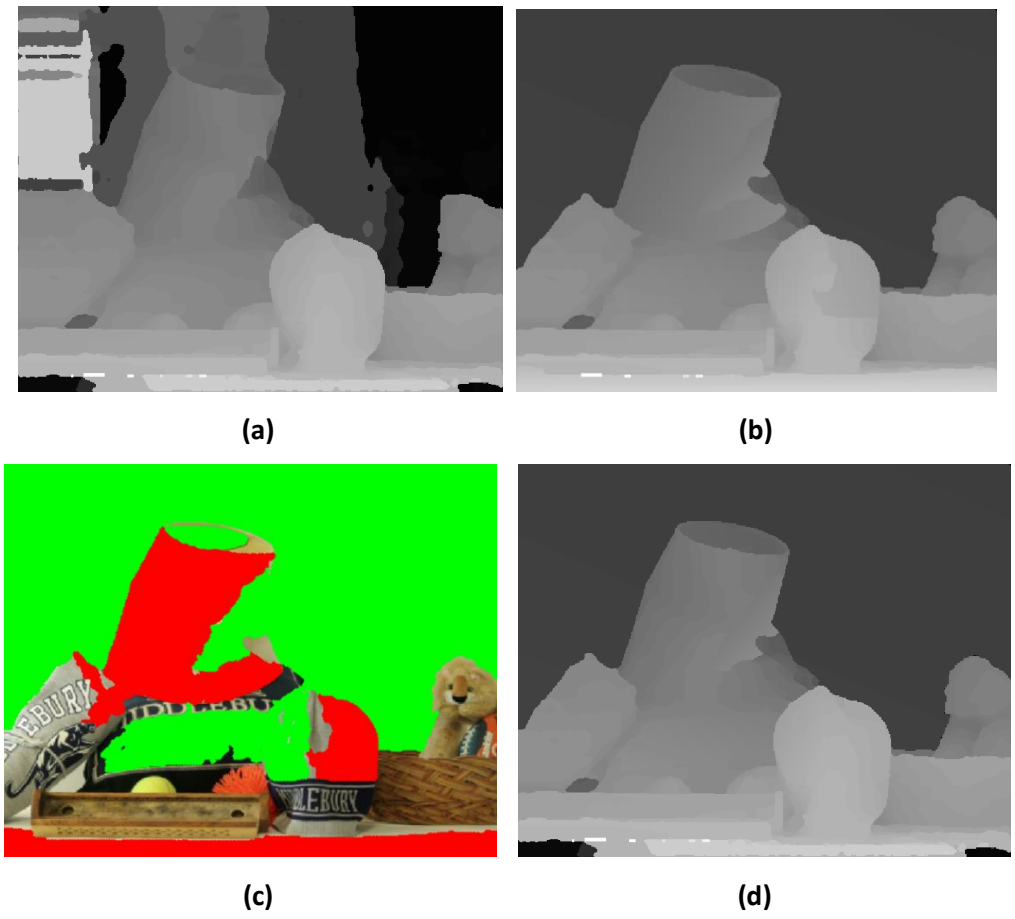
5.2.4.4 Planar fitting

Afterwards, the reliable pixels and their disparities (red points in Figure 30a) are used to fit a planar surface to the segment. The robust method of Principal Components Analysis (PCA) described in [70] is used to estimate the parameters of the plane. The two first principal components define the plane. Let the estimated plane be: $d_p(\mathbf{x}) = \mathbf{p}^T \cdot \mathbf{x}$, where $\mathbf{p}=[p_1, p_2]^T$. Then each $\mathbf{x} \in S$ is assigned the disparity $d_p(\mathbf{x})$. The new disparity values inside the segment are depicted with blue in Figure 30a.

A second example of uniform area handling is given considering the Cones stereo pair. In brief, Figure 28c shows the overall outliers map. For the mean-shift segment of Figure 28d (marked with blue colour), the histogram of the disparities of the inlier pixels inside this segment is depicted in Figure 29b. The reliable subset of inlier pixels includes the pixels whose disparities belong to the 1st histogram part of

Figure 29b. The reliable pixels and their disparities (red points in Figure 30b) are used to fit a planar surface to the segment. The new disparity values inside the segment are depicted with blue in Figure 30b.

Figure 31 shows three examples of uniform region handling. In the first and second rows of Figure 31, the disparity results before and after uniform regions handling are visualized, respectively. The first and second columns include the result of handling the blue-colored segments of Figure 28b and Figure 28d, respectively. The third column shows an example for the Teddy stereo pair. The examples in the second and third column show clearly the improvements in the disparity maps after applying the plane fitting process.



(a) **(b)**

(c) **(d)**

Figure 32. Illustration of: (a) the disparity map without uniform areas handling, (b) the disparity map with uniform areas handling (without exploiting the MED_{fit} verification metric), (c) the uniform areas, which are denoted with green, where the disparity plane fitting is assumed as successful according to $MED_{fit} < 0.5$, (d) the disparity map after uniform area handling for the areas that satisfy $MED_{fit} < 0.5$.

However, it is not always valid to assume that large areas with low texture are planar. Additionally, some large areas may have been wrongly segmented, leading to

inaccurate plane fitting. Therefore, a specific metric is adopted, which is used to verify if the planar fitting is successful. This metric is the median of the absolute differences between the initial disparities of the reliable pixels and the disparities of the reliable pixels that are estimated after the plane fitting and is defined as: MED_{fit} (measured in disparity units). The condition $MED_{fit} < 0.5$ has to be satisfied, in order to consider the planar fitting as successful.

Figure 32 visualizes an example of uniform areas handling applied on the Midd1 stereo pair, which belongs to the extended stereo dataset [71] (see section 6.1) and contains large low-textured areas. In Figure 32a the estimated disparity map without applying the uniform areas handling is depicted. It is obvious that disparity estimation is not reliable in low-textured areas. Figure 32b shows the disparity map after applying uniform areas handling to all low-textured areas. Figure 32c visualizes with green the low-textured areas with $MED_{fit} < 0.5$ and with red the low-textured areas with $MED_{fit} \geq 0.5$. In Figure 32d the disparity map after applying uniform areas handling only for the green low-textured areas is depicted. The disparity error for the case of all regions and $\Delta d > 1$ for the Midd1 stereo pair is 40.65%, 14.88% and 9.69% for the disparity maps of Figure 32a, Figure 32b and Figure 32d, respectively. Therefore, this example verifies the efficiency of the uniform areas handling to decrease the disparity estimation error.

A median filter using a 5x5 neighborhood is applied to the disparity result that is generated after executing all the steps of methodology A, in order to remove spurious disparities before acquiring the final disparity map.

5.3 Summary

The current chapter presented the disparity optimization and disparity refinement steps for methodologies A and B.

The disparity optimization for both methodologies is based on the semi-global optimization approach, where two novel ideas are introduced to improve its performance. The first idea, which is used by methodology A, concerns the introduction of a new criterion that relies on mean-shift segmentation for the

detection of depth discontinuities. This criterion, which is used in the definition of the smoother penalty terms, checks whether or not two neighboring pixels, along a path direction, belong to the same mean-shift segment. The second idea, which is used by methodology B, concerns the introduction of a weighted variant of the semi-global optimization, where the path costs of a considered pixel may have different weights depending on the number of the pixels that precede the considered pixel along each path direction. Practically, for a considered pixel, the path direction that contains much more non-outlier (inlier) pixels than the other directions will receive higher weight.

The disparity refinement step in methodology A comprises outliers handling, disparity edges refinement and uniform areas handling.

Outliers handling in methodology A, is performed by combining the “Basic outlier handling’’ scheme and the “Mean-shift segmentation-based outlier handling” scheme. The ``Basic outlier handling’’ scheme sets the disparity of an outlier pixel equal to the minimum disparity between the disparities of its spatially closest inlier pixels on its left and its right side. On the other hand, the ``Mean-shift segmentation-based outlier handling” scheme initially classifies segments into reliable and unreliable segments. For each outlier pixel inside a reliable segment, a voting scheme that counts the disparities of the inlier pixels that belong to the same segment as the outlier pixel is applied. The most repeated disparity is propagated to the outlier pixel. For an unreliable segment, the reliable neighboring segment that will propagate its prevalent disparity to this unreliable segment is the one that has the most similar colour to the unreliable segment. Then, methodology A applies a two-step approach to perform disparity edges refinement. The first step handles the disparity edges at a coarser level and the second one at a finer level. Uniforms areas handling encompasses disparity histogram analysis, which helps to acquire a reliable subset of inlier pixels that can be used to perform accurate disparity plane fitting on these uniform areas.

The disparity refinement step in methodology B comprises outliers handling and disparity edges refinement.

Outliers handling in methodology B is performed by applying sequentially “Generic outliers handling” and “Background outliers handling”. “Generic outliers

handling" generates two disparity histograms for the left and the right side of an outlier pixel. If either of two specific conditions is met, then the disparity of the outlier pixel is set equal to the disparity which corresponds to the bin containing the maximum value in the disparity histogram of the left or the right side. `` Background outliers handling'' sets the disparity of an outlier pixel equal to the minimum disparity between the disparities of its spatially closest inlier pixels on its left and its right side. The filled regions are then smoothed using a bilateral filter. After outliers handling, methodology B applies a straightforward efficient approach to refine the disparity estimation at the edges. This approach builds a disparity histogram for each pixel lying on a disparity edge. The disparity of the edge pixel is set equal to the disparity that corresponds to the bin containing the maximum value in the generated disparity histogram.

The next chapter provides the evaluation and the experiments that were conducted to test the performance of methodology A and methodology B.

Chapter 6

6 Evaluation and experiments

Chapter 6 presents the experimental evaluation of the proposed methodologies.

6.1 Datasets

The experiments are performed using Middlebury stereo pairs that belong to three datasets. The four stereo image pairs of the Middlebury online stereo benchmark dataset are used to present computational time results, evaluate the accuracy of the presented methodologies and select optimum parameters. The online dataset includes the Tsukuba, Venus, Teddy and Cones stereo pairs. The Tsukuba pair was used to demonstrate methodology A, while the Teddy pair was used to demonstrate methodology B. The left views of these four image pairs and their ground truth disparity maps are referred as the Dataset 2003 in Figure 33.

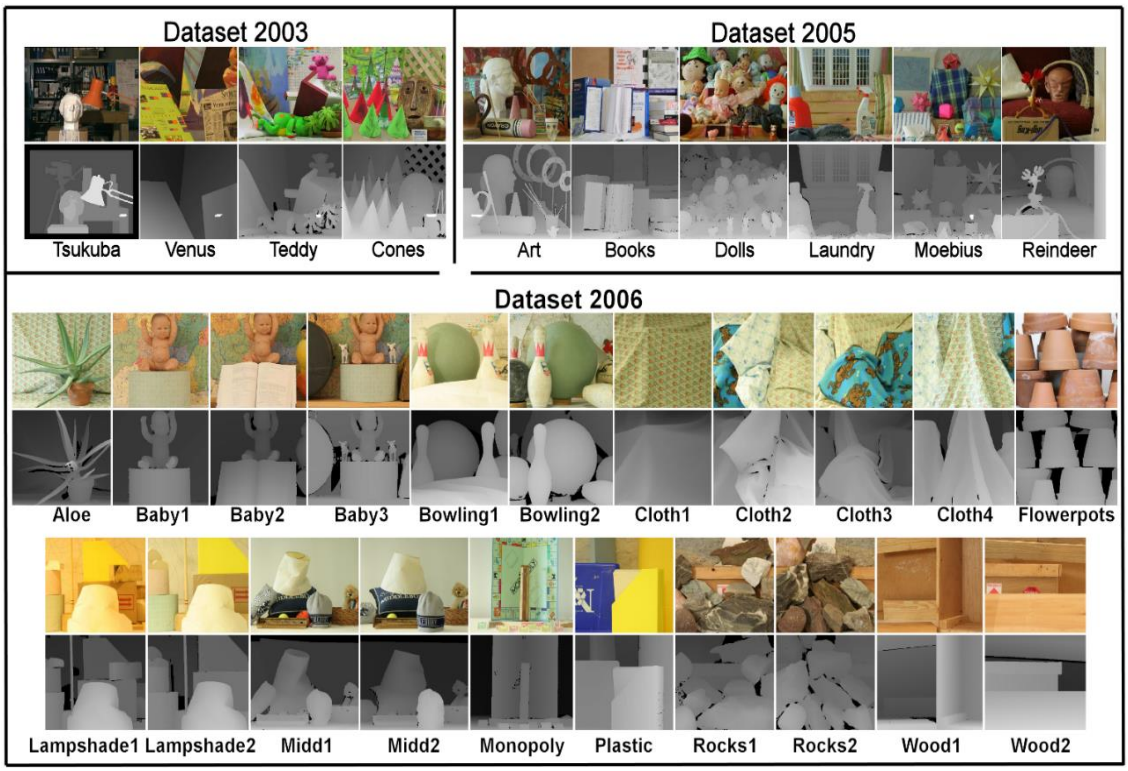


Figure 33. Left views of the stereo image pairs and their corresponding ground truth disparity map

Two more datasets, which include in total twenty seven stereo pairs, have been also used for evaluating the accuracy of the presented methodologies. The left views and the corresponding ground truth disparity maps from the six pairs of the Dataset 2005 and the twenty one pairs of the Dataset 2006 are visualized in Figure 33.

6.2 Computational analysis

6.2.1 Computational analysis for methodology A

A non-optimized C++ implementation of methodology A is used to report on the required computational time. The algorithm was executed on a desktop PC with Core i7-3770 3.40 GHz CPU and 8 GB RAM. The total processing time, using as input the four stereo pairs of the Middlebury evaluation benchmark [68], is indicated in Table 2. The measured time is the average of 5 separate runs. Additionally, this table provides the percentage of the total time that is spent for each of algorithm's steps, which include: 1. the Matching Cost Computation step (M.C.) (subsection 4.1.1), 2.the Cost Aggregation step (C.A.) (subsection 4.2.1), 3. the Disparity Optimization step (D.O) (subsection 5.1.2) and 4. the Disparity Refinement step (D.R.) (subsections 5.2.2.1, 5.2.3.1 and 5.2.4). The Cost Aggregation is the most computational expensive step (on average 91.84 % of the total processing time). Nevertheless, this step can be parallelized since cost aggregation can be performed independently for non-overlapping parts of the image.

Image	Resolution	Disp.Levels	Meth. A (s)	M.C.(%)	C.A.(%)	D.O.(%)	D.R.(%)
Tsukuba	384 x 288	15	24.33	4.30	88.93	0.93	5.84
Venus	434 x 383	20	49.25	3.71	89.72	0.92	5.65
Teddy	450 x 375	60	154.23	2.89	94.35	0.86	1.90
Cones	450 x 375	60	154.78	2.82	94.45	0.92	1.81

Table 2. Computational time in seconds and the percentage of time spent on each step of methodology A.

Concluding, most parts of the algorithm have low computational cost. The step of the algorithm with increased computational cost includes the adaptive support

weight cost aggregation (see subsection 4.2.1). However, this time consuming part can be implemented in Graphics Processing Units (GPU) as can be verified in [28]. Additionally, there are works, such as [69], [72] that propose approximations to derive fast implementations of the original adaptive support weight algorithm [21]. The drawback of these methods is that they sacrifice quality for high computational speed [28].

6.2.2 Computational analysis for methodology B

A C++ implementation of the methodology B is used to report on the required computational time. The algorithm was executed on a desktop PC with Core i7-3770 3.40 GHZ CPU and 8 GB RAM. The low processing time using as input each of the four stereo pairs of the Middlebury evaluation benchmark [68] is indicated in Table 3. The measured time is the average of 5 separate runs. Table 3 provides the percentage of the total time that is spent for: 1. the Matching Cost Computation step (M.C.) (subsection 4.1.2), 2. the Cost Aggregation step (C.A.) (subsection 4.2.2), 3. the Disparity Optimization step (D.O) (subsection 5.1.2) and 4. the Disparity Refinement step (D.R.) (subsections 5.2.2.2 and 5.2.3.2). The Cost Aggregation step, which relies on content-based guided filtering, is the most computational expensive step (on average 61.78 % of the total processing time).

Image	Resolution	Disp.Levels	Meth. B (s)	M.C.(%)	C.A.(%)	D.O.(%)	D.R.(%)
Tsukuba	384 x 288	15	1.9	20.45	59.02	15.90	4.63
Venus	434 x 383	20	3.6	20.95	61.06	15.32	2.67
Teddy	450 x 375	60	9.7	19.91	63.61	13.34	3.14
Cones	450 x 375	60	9.6	19.63	63.42	13.82	3.13

Table 3. Computational time in seconds and the percentage of time spent on each step of methodology B.

Concluding, the step of the methodology B with increased computational cost includes the content-based guide image filtering (see subsection 4.2.2). However, this part can be implemented in GPU as can be verified in [18], [27]. The semi-global

optimization method, which is used in the Disparity Optimization step, can be also implemented in GPU according to [16], [73]. Therefore, methodology B is appropriate for real-time GPU implementation.

6.3 Parameters selection

6.3.1 Set of optimum parameters for methodology A

The parameters used for the experiments are the same for all tested stereo pairs. More specifically, β is set equal to $\beta = 0.3$, while the parameters used for the cost functions are $\lambda_{\text{RGB}} = 30$, $\lambda_{\text{CEN}} = 45$ and $\lambda_{\text{SIFT}} = 45$ (see subsection 4.1.1). The radius of the support area (see subsection 4.2.1.1) is set equal to $R_{\text{S}} = 19$ and the adaptive weight parameters are $\gamma_{\text{c}} = 8$ and $\gamma_e = R_S$. The parameters used in subsection 5.1.2 are $\Pi_1 = 0.2, \Pi_2 = 0.6$ and $\tau_{\text{so}} = 10$.

	Best	*β=0.25*	*β=0.35*	λ_{RGB}=25	λ_{RGB}=35	γ_c=7	γ_c=9	λ_{CEN}=40	λ_{CEN}=50	R_S=17	R_S=21	**No Crit.**
Avg. Rank	15.6	17.4	15.9	16.5	17.5	18.2	17.2	17.7	17.1	16.2	17.4	18.5
Nonocc (%)	2.08	2.11	2.09	2.10	2.11	2.10	2.10	2.12	2.10	2.09	2.10	2.16
All (%)	4.51	4.57	4.52	4.51	4.53	4.53	4.54	4.54	4.53	4.52	4.54	4.57
Disc (%)	6.41	6.41	6.43	6.43	6.41	6.51	6.43	6.41	6.48	6.46	6.49	6.39

Table 4. Parameters testing for methodology A.

The column ``Best'' of Table 4 gives, for methodology A, the numeric results from the Middlebury Stereo evaluation for the disparity maps extracted using the optimum parameters. The results include the overall performance measure (``Avg. Rank''), the error in non-occluded regions (``Nonocc''), the error in all regions (``All'') and the error near depth discontinuities (``Disc''). In subsection 6.5.1.4 further parameters testing is performed.

6.3.2 Set of optimum parameters for methodology B

The parameters used for the experiments are the same for all tested stereo pairs. More specifically, the parameters used for the estimation of the cost term (see

subsection 4.1.2) are defined as: $\{\alpha_1,\alpha_2,T_{\text{gra}},T_{\text{gab}},T_{\text{BT}}\}=\{0.75,0.20,2/255,4/255,7/255\}$. The variables used for the cost filtering are the smoothness parameter ε (see subsection 4.2.2.1), which is set to $\varepsilon=0.0001$ and the parameter R_{S} that defines the size of the rectangular window (see subsection 4.2.2.2), which is set to $R_{\text{S}}=17$. The selection of $R_{\text{S}}=17$ is based on the experiments described in subsection 6.5.2.2. The parameters used in subsection 5.1.2 are $\Pi_1=0.002$, $\Pi_2=0.006$ and $\tau_{\text{so}}=10/255$. Finally, the colour difference threshold τ_1 in subsection 5.2.2.2 is set to $\tau_1=10/255$.

	"Best"	"Adapt. Windows" as in [27]	"Simple S-G" as in [44]
Avg. Rank	16.8	23.4	18.0
Nonocc (%)	1.91	2.12	1.95
All (%)	4.68	4.85	4.70
Disc (%)	6.41	6.41	6.43

Table 5. Evaluation results for methodology B.

The column ``Best'' of Table 5 gives, for methodology B, the numeric results from the Middlebury Stereo evaluation for the disparity maps extracted using the optimum parameters. The results include the overall performance measure (``Avg. Rank''), the error in non-occluded regions (``Nonocc''), the error in all regions (``All''), the error near depth discontinuities (``Disc'') and the average percent of bad pixels (``APBP'').

6.4 Disparity results

6.4.1 Disparity results of methodology A

The disparity results of methodology A, for the optimum parameters set, accompanied with the disparity error maps as extracted by the Middlebury evaluation system are visualized in Figure 34. Errors in non-occluded and occluded regions are marked in black and gray respectively in the second row of Figure 34.

Algorithm	Avg. Rank	Tsukuba			Venus			Teddy			Cones		
		nonocc	all	disc	nonocc	all	disc	nonocc	all	disc	nonocc	all	disc
IGSM[74]	10.5	0.93	1.37	5.05	0.07	0.17	1.04	4.08	5.98	11.4	2.14	6.97	6.27
TSGO[75]	13.8	0.87	1.13	4.66	0.11	0.24	1.47	5.61	8.09	13.8	1.67	6.16	4.95
JSOSP-GCP[76]	15.2	0.74	1.34	3.98	0.08	0.16	1.15	3.96	10.1	11.8	2.28	7.91	6.74
Methodol. A	15.6	1.02	1.23	5.51	0.08	0.20	1.11	5.16	9.43	13.0	2.07	7.16	5.97
Methodol. B	16.8	1.01	1.32	5.17	0.08	0.21	1.17	4.35	9.83	12.3	2.19	7.35	6.43
SSCBP[77]	18.2	1.05	1.39	5.57	0.10	0.16	1.39	3.44	8.32	9.95	2.60	7.13	7.23
ADCensus[16]	18.8	1.07	1.48	5.73	0.09	0.25	1.15	4.10	6.22	10.9	2.42	7.25	6.95

Table 6. The rankings in the Middlebury benchmark.

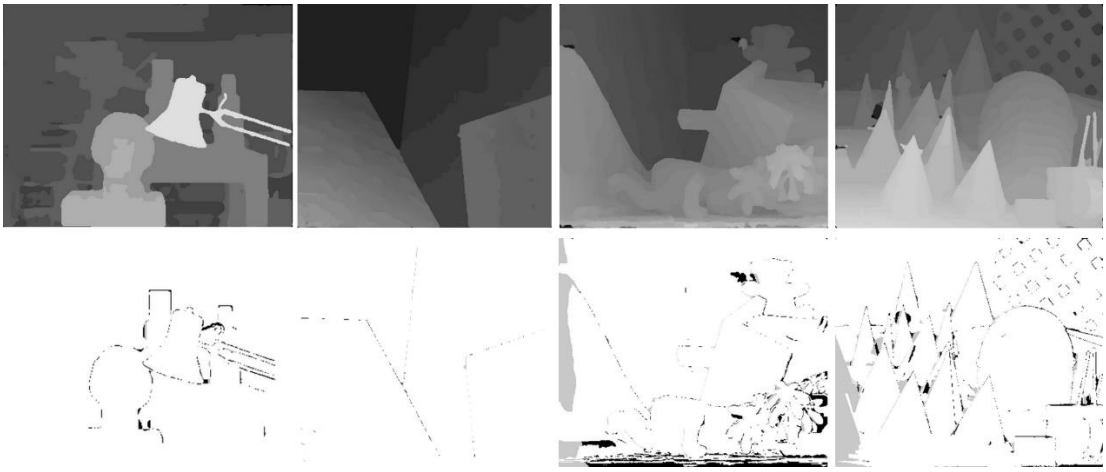


Figure 34. Disparity maps generated using methodology A and their corresponding disparity error maps for error threshold 1.

The ranking results in Table 6, for error threshold equal to 1, indicate that methodology A is 4th out of 164 methods that are included in the Middlebury Stereo Evaluation. However, no information on the 1st [74] ranked methods is available, since it is currently under review. Therefore, methodology A ranks 3rd among already published methods. More specifically, the proposed method ranks: 10th for the ``Tsukuba'' image pair, 4$^{th}$ for the Venus image pair, 37$^{th}$ for the Teddy image pair and 4$^{th}$ for the ``Cones'' image pair.

The 37th position in the ranking for the Teddy image pair is because of the very slanted surface at the bottom of the image, where the proposed method cannot handle well the very slanted surface. However, it can be deduced from the experimental results that the proposed method outperforms the rest of the published

stereo methods, which are evaluated online in the Middlebury stereo evaluation benchmark, in image areas excluding very slanted surfaces.

6.4.2 Disparity results of methodology B

The disparity results of methodology B, for the optimum parameters set, accompanied with the disparity error maps are visualized in Figure 35.

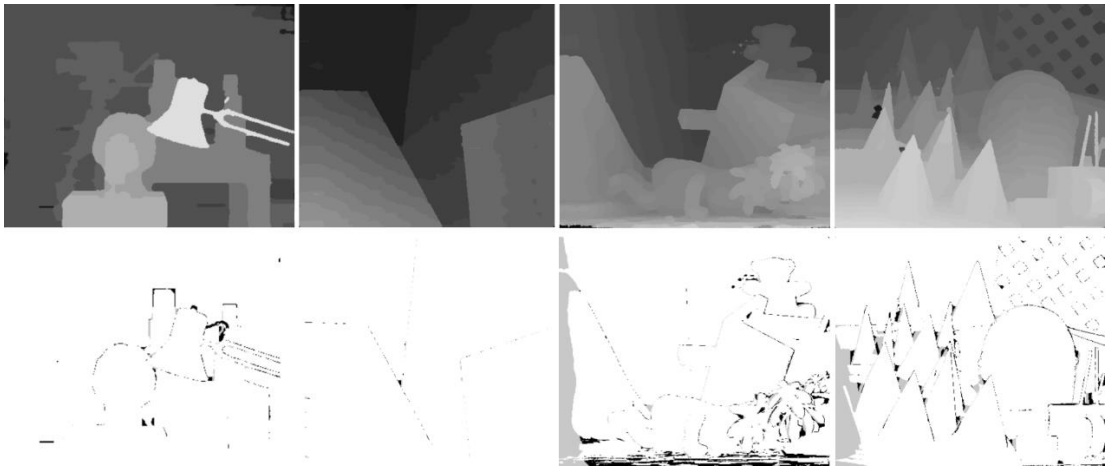


Figure 35. Disparity maps generated using methodology B and their corresponding disparity error maps for error threshold 1.

The ranking results in Table 6, for error threshold equal to 1, indicate that methodology B is 5th out of 164 methods that are included in the Middlebury Stereo Evaluation. Since the 1st [74] ranked methods is currently under review, methodology B ranks 4th among already published methods (methodology A is included in the published methods that rank above methodology B). This is an important achievement bearing in mind the reduced computational complexity of this algorithm and its suitability to be implemented in GPU. Moreover, though methodology B is less accurate than methodology A and the JSOSP-GCP approach [76], it is faster than them. This fact is evident in Table 7 that displays the computational times of methodology A, methodology B and the approach in JSOSP-GCP (the computational times regarding JSOSP-GCP were obtained from [76] after converting minutes into seconds). Additionally, methodology B outperforms in terms of disparity estimation accuracy the methods presented in [18] and [27], which also exploit the guided image filter and rank in positions 48 and 18, respectively.

In more detail, regarding the Middlebury Stereo Evaluation, the proposed method ranks: 12th for the ``Tsukuba'' image pair, 6th for the Venus image pair, 29th

for the Teddy image pair and 15th for the ``Cones'' image pair.

Image	Resolution	Disp.Levels	Meth. A	Meth. B	JSOSP-GCP[76]
Tsukuba	384 x 288	15	24.33	1.9	143.4
Venus	434 x 383	20	49.25	3.6	249.0
Teddy	450 x 375	60	154.23	9.7	262.8
Cones	450 x 375	60	154.78	9.6	306.6

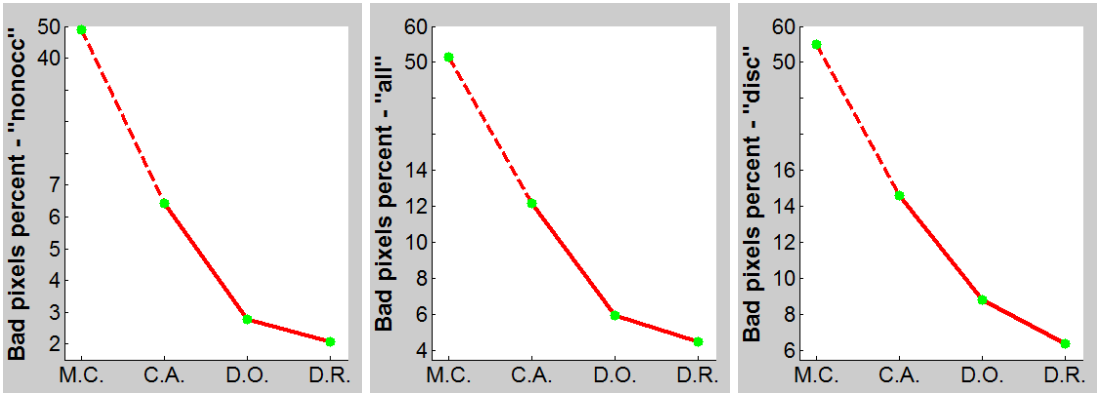
Table 7. Comparison of computational times in seconds.

6.5 Evaluation Results

6.5.1 Evaluation Results of methodology A

6.5.1.1 Evaluation of methodology A

The initial disparity map (see Figure 9) that is generated after applying WTA to the cost volume $C_{\text{R-C}}(\mathbf{x},\mathbf{d})$ ($C_{\text{R-C}}(\mathbf{x},\mathbf{d})$ has been computed in subsection 4.1.1.3 via Equation (8)) is heavily corrupted with noisy disparities. This subsection examines how the accuracy of the initial disparity map, which has been estimated via 1. the Matching Cost Computation step (M.C.) , is improved after applying sequentially: 2. the Cost Aggregation step (C.A.) (subsection 4.2.1), 3. the Disparity Optimization step (D.O) (subsection 5.1.2) and 4. the Disparity Refinement step (D.R.) (subsections 5.2.2.1, 5.2.3.1 and 5.2.4).



(a) **(b)** **(c)**

Figure 36. Average percent of bad pixels after applying sequentially the steps of methodoylogy A for (a) non-occluded regions, (b) all regions and (c) near depth discontinuities regions.

Figure 36 depicts the average percent of bad pixels for the disparity maps of the four Middlebury image pairs, which are generated after applying sequentially each step of methodology A. This table includes results for non-occluded regions (``Nonocc''), all regions (``All'') and regions near depth discontinuities (``Disc'').

The cost aggregation step significantly enhances the initial disparity map. This is evident in the results of Figure 36 where the average percent of bad pixels drastically reduces from the Matching Cost Computation (M.C.) step to the Cost Aggregation (C.A.) step. The disparity map accuracy is further improved after applying disparity optimization. This is evident in Figure 36, where the percent of bad pixels declines from the Cost Aggregation (C.A.) step to the Disparity Optimization (D.O.) step. Finally, the Disparity Refinement (D.R.) step helps in further lowering the percent of bad pixels with respect to the Disparity Optimization (D.O.) step, as it is also shown in Figure 36.

The improvement in the disparity map quality, introduced by the aforementioned steps, is also visually demonstrated. As it is obvious in Figure 9, the disparity map, which is acquired via the matching cost computation (M.C.) step, is severely corrupted with estimation-error noise. After applying the cost aggregation (C.A.) step the noise is removed, as it is evident in Figure 14a. The disparity optimization step (D.O.) further improves the disparity results. This is clearly seen observed from the comparison between Figure 14a and Figure 20a. The disparity map, which is given on the left image in the 2nd row of Figure 31, shows the disparity map after performing the disparity refinement (D.R.) step to the disparity map of Figure 20a. The outlier regions of the disparity map in Figure 20a have been efficiently handled in the disparity map given on the left image in the 2nd row of Figure 31.

In the following, particular evaluations of several steps of methodology A are also provided.

6.5.1.2 Evaluation of the two-phase combination strategy

This subsection provides the evaluation of the two-phase combination strategy, which is part of the cost aggregation step. The improvement in the accuracy

of the disparity map resulting via WTA from $V_{\text{R-C}}^{'}$, which is achieved by using the two-phase combination strategy of subsection 4.2.1.2, is evaluated according to the Middlebury online evaluation system. Table 8 depicts the average percent of bad pixels for the disparity maps generated using the four Middlebury image pairs. This table includes results for non-occluded regions (``Nonocc''), all regions (``All'') and regions near depth discontinuities (``Disc'').

	Init.	**Phase1**	**Phase2**	**CENSUS**	**SIFT**
Nonocc (%)	8.81	7.91	6.43	18.5	15.0
All (%)	14.4	13.6	12.2	23.1	19.8
Disc (%)	15.9	15.6	14.6	27.3	27.8

Table 8. Evaluation of the two-phase combination strategy of methodology A.

The evaluation results for the disparity map resulting via WTA from $V_{\text{R-C}}^{'}$ are given in the ``Init.'' column. The evaluation results for the disparity map resulting via WTA from $V_{\text{R-C}}^{'}$ (which is acquired after applying first combination phase) are given in the ``Phase1'' column. The average numeric results for the disparity map d_{LR} resulting via WTA from $C_f(\mathbf{x},\mathbf{d})$ (which is acquired after applying second combination phase) are given in the ``Phase2'' column. Obviously, each combination phase assists in improving the accuracy of the generated disparity map.

Additionally, Table 8 includes in the ``CENSUS'' column the evaluation results for the disparity map resulting via WTA from $V_{\text{CEN}}$ and in the ``SIFT'' column the evaluation results for the disparity map resulting via WTA from V_{SIFT}. Though, the results in ``CENSUS'' and ``SIFT'' are worse that the results in ``Init.'' the efficient exploitation of V_{CEN} and V_{SIFT} in the two-phase combination strategy improves the disparity estimation accuracy.

6.5.1.3 Evaluation of the disparity refinement process

Furthermore, the Middlebury online benchmark is exploited in order to

examine the improvement introduced by the proposed disparity refinement steps. Figure 37 depicts how the average percent of bad pixels decreases after applying sequentially each of the disparity refinement steps, which include outlier handling, disparity edges refinement and uniform regions handling. Figure 37 includes results for non-occluded regions (see Figure 37a), all regions (see Figure 37b) and regions near depth discontinuities (see Figure 37c). As it is expected, the outlier handling decreases the bad pixels percent more than the rest refinement steps, since it handles large outlier areas. Disparity edges refinement and uniform regions handling improve further the accuracy, so that the proposed framework becomes the top ranked published method in the Middlebury stereo evaluation.

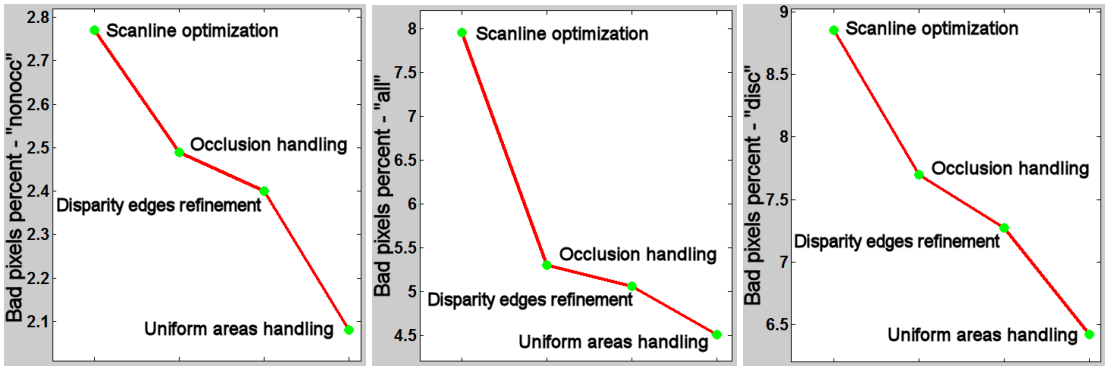


Figure 37. Average percent of bad pixels after applying sequentially refinement steps for (a) nonoccluded regions, (b) all regions and (c) near depth discontinuities regions.

6.5.1.4 Further parameters testing

As mentioned in subsection 6.3.1, the column ``Best'' of Table 4 gives the numeric disparity estimation results using optimum parameters. In the rest columns of Table 4, the results in the case that all parameters are kept the same as the optimum ones, except for the parameter in the top of the column, are provided. For each parameter a smaller and a larger value than the optimum one are tested. Table 4 verifies that the optimum parameters give the best results.

The last column of Table 4, with the annotation ``No criterion'', gives the results of this method for the best set of parameters, with the difference that in this case the introduced second criterion for the definition of the smoothness terms in Equation (34) is not used. The results prove that without the exploitation of the

second criterion the disparity accuracy decreases.

The segmentation maps are exploited in different stages of this method. Therefore, it is important to verify that small variations to the optimum parameters $(\sigma_s, \sigma_r) = (3,3)$ that adjust the segmentation result (see subsection 3.4.3) do not affect significantly the performance of this method. Table 9 exhibits the error results for different values of the spatial radius and space feature radius. The rest of parameters are set to their optimum value.

	$(\sigma_s,\sigma_r)=(2,3)$	$(\sigma_s,\sigma_r)=(3,4)$	$(\sigma_s,\sigma_r)=(4,4)$
Avg. Rank	18.5	19.4	18.6
Nonocc (%)	2.16	2.14	2.12
All (%)	4.77	4.69	4.69
Disc (%)	6.51	6.54	6.59

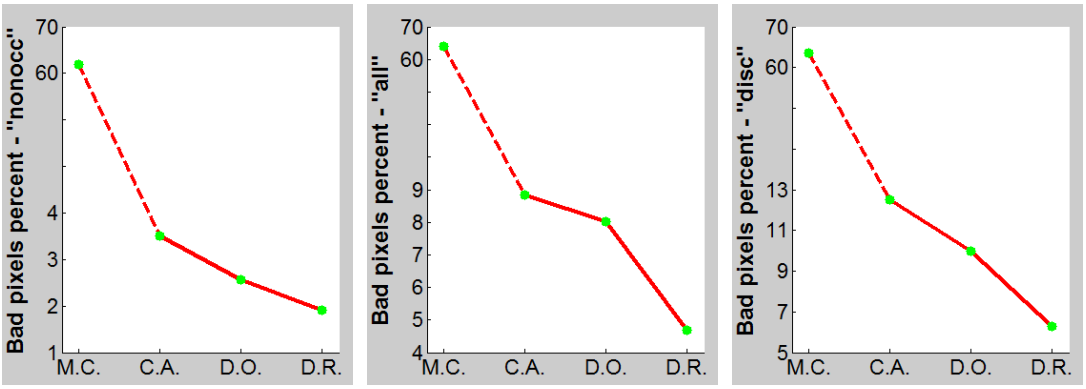
Table 9. Segmentation parameters testing for methodology A.

For all parameter tests, the proposed method ranks in the top five ranking positions though the disparity accuracy decreases. This fact proves that this approach maintains its good disparity estimation accuracy even with changes to the optimum parameters.

6.5.2 Evaluation Results of methodology B

6.5.2.1 Evaluation of methodology B

The initial disparity map (see Figure 10) that is generated after applying WTA to the cost volume $C(\mathbf{x},\mathbf{d})$ ($C(\mathbf{x},\mathbf{d})$ has been computed in subsection 4.1.2 via Equation (18)) is heavily corrupted with noisy disparities. This subsection examines how the accuracy of the initial disparity map, which has been estimated via 1. the Matching Cost Computation step (M.C.) (subsection 4.1.2), is improved after applying sequentially: 2. the Cost Aggregation step (C.A.) (subsection 4.2.2), 3. the Disparity Optimization step (D.O) (subsection 5.1.2) and 4. the Disparity Refinement step (D.R.) (subsections 5.2.2.2 and 5.2.3.2).



(a) **(b)** **(c)**

Figure 38. Average percent of bad pixels after applying sequentially the steps of methodoylogy A for (a) non-occluded regions, (b) all regions and (c) near depth discontinuities regions.

The cost aggregation step significantly enhances the initial disparity map. This is evident in the results of Figure 38, where the average percent of bad pixels significantly declines from the Matching Cost Computation (M.C.) step to the Cost Aggregation (C.A.) step. The disparity map accuracy is further improved after applying disparity optimization. This is evident in Figure 38, where the percent of bad pixels reduces from the Cost Aggregation (C.A.) step to the Disparity Optimization (D.O.) step. The improvement is stronger for the regions near depth discontinuities. Finally, the Disparity Refinement (D.R.) step helps in further lowering the percent of bad pixels with respect to the Disparity Optimization (D.O.) step, as it is also shown in Figure 38.

The improvement in the disparity map quality, introduced by the aforementioned steps, is also visually demonstrated. The disparity map, which is estimated via the matching cost computation (M.C.) step, is heavily corrupted with estimation-error noise, as it is evident in Figure 10. After applying the cost aggregation (C.A.) step the noise is eliminated, as it is evident in Figure 16a. The disparity optimization step (D.O.) further improves the disparity results. This fact is visually verified from the comparison between Figure 16a and Figure 21a. Figure 27b shows the disparity map after performing the disparity refinement (D.R.) step to the disparity map of Figure 21a. The outlier regions of the disparity map in Figure 21a have been efficiently handled in the disparity map in Figure 27b.

In the following, particular evaluations for several steps of methodology B are also provided.

6.5.2.2 Experiments on the definition of support windows

In order to prove why the exploitation of rectangular support windows of two sizes (as suggested in subsection 4.2.2.2) enhances the disparity estimation results, experiments using support windows of either rectangular or square shape have been performed.

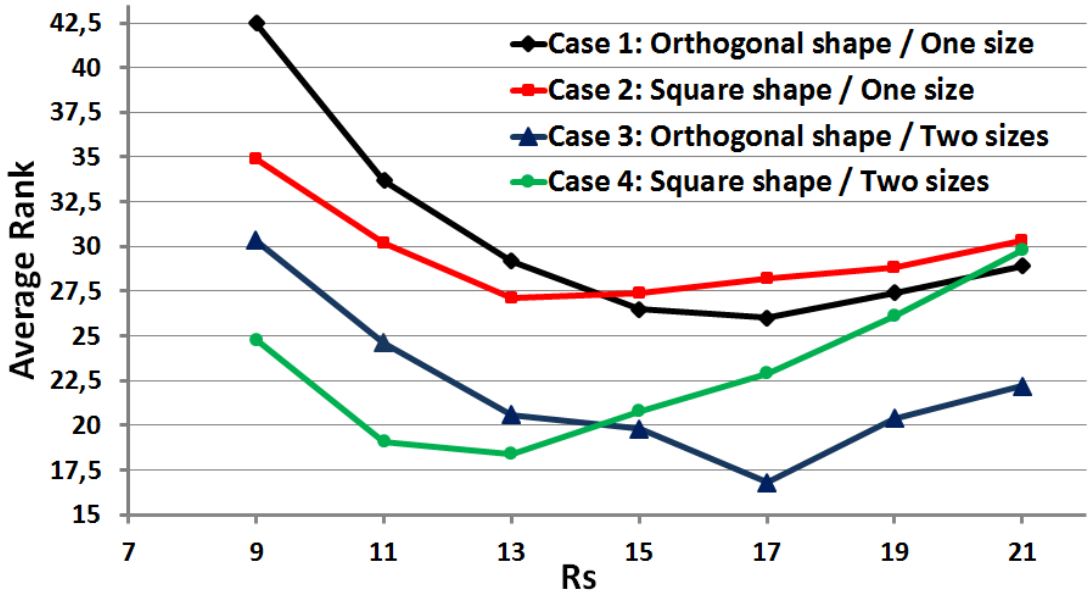


Figure 39. Average Rank against R_S for four different cases of defining support windows sizes.

In specific, the following cases of defining support windows have been examined:

- Case 1: Use one rectangular support window with size R_S x$\lceil R_S / 2 \rceil$.
- Case 2: Use one square support window with size R_S x R_S .
- Case 3: Use two rectangular support windows with sizes R_S x$\lceil R_S / 2 \rceil$ and $2R_S$ x R_S .
- Case 4: Use two square support windows with sizes R_S x R_S and $2R_S$ x R_S .

Those four cases affect subsection 4.2.2.2. In more detail, when using just one support window (``Case 1'' or ``Case 2'') the selection of the appropriate support window size for each pixel is not required, while when using two support windows

(``Case 3'' or ``Case 4'') the condition ``If $\overline{\mathrm{M}}(\mathbf{p}) > R_{\mathrm{S}}$ '' (see subsection 4.2.2.2) is examined to decide which of the two windows is more appropriate to determine the filtered cost of each pixel.

The curves in Figure 39 show the Average Rank (as estimated according to the online Middlebury evaluation) for each of the above four cases, for different values of R_{S} . An important finding is that between ``Case 1'' and ``Case 2'', ``Case 1'' (rectangular support window) gives better Average Rank than ``Case 2'' (square support window). Moreover, by comparing ``Case 1'' and ``Case 2'' with ``Case 3'' and ``Case 4'', it is evident that the use of two support windows sizes gives a better Average Rank. Finally, it is shown that ``Case 3'' (this is the case proposed in subsection 4.2.2.2) gives the best disparity estimation results among all cases. The value of R_{S} for which the best Average Rank is accomplished is $R_{\mathrm{S}} = 17$.

6.5.2.3 Further parameters testing

As mentioned in subsection 6.3.2, the column ``Best'' of Table 5 gives the numeric disparity estimation results using methodology B with the optimum parameters. In the rest of the columns of Table 5, experimental results after making some modifications to the methodology are demonstrated.

In order to evaluate how the proposed improvement with respect to the semi-global optimization (this improvement includes the weighted average of path costs according to section 5.1.2) ameliorates the disparity results, numeric results for the case where the semi-global approaches of [16], [44] have been used, instead of the weighed semi-global approach (which is proposed in 5.1.2), have been included. The approaches in [16], [44] use a simple average of path costs, while the threshold used for the identification of depth discontinuities is constant. The numeric disparity results, which have been estimated using the simple average of path costs as in [44] are given in the column of Table 5, with the annotation ``Simple S-G''. Except for the semi-global optimization step the other steps of methodology B are applied as they are. The differences between the column ``Best'' and ``Simple S-G'' prove that without the weighted semi-global optimization the disparity estimation accuracy decreases.

	$(\sigma_s,\sigma_r)=(2,3)$	$(\sigma_s,\sigma_r)=(3,4)$	$(\sigma_s,\sigma_r)=(4,4)$
Avg. Rank	18.0	16.9	17.5
Nonocc (%)	1.94	1.91	1.90
All (%)	4.71	4.68	4.69
Disc (%)	6.31	6.27	6.26

Table 10. Segmentation parameters testing for methodology B.

The mean-shift segmentation map (subsection 3.4.3) is exploited for selecting the appropriate support window size of each pixel (see subsection 4.2.2.2). Therefore, it is important to verify that small variations to the optimum parameters $(\sigma_s,\sigma_r)=(3,3)$ that adjust the segmentation result do not affect significantly the performance of this method. Table 10 shows the error results for different values of the spatial radius and space feature radius. For the pairs of $(\sigma_s,\sigma_r)=(2,3)$, $(\sigma_s,\sigma_r)=(3,4)$ and $(\sigma_s,\sigma_r)=(4,4)$ this methodology remains in the third position. Hence, it is deduced that even varying the segmentation parameters the method remains in the top performers.

6.5.2.4 Comparison with the related approach of [27]

A relevant work that uses adaptive guided image filtering is presented in [27]. However, there are significant differences between [27] and methodology B, regarding the selection of the support window for each pixel. In [27] the support window for each pixel is based on a skeleton that is built from four arms stretching in four directions, where the borders of the support window are determined directly by the endpoints of the arms. Therefore, for each pixel there is a different support window. On the contrary, in methodology B two support window sizes are used.

A main advantage of the guided filter is that the computation cost is independent to the size of the selected support window. This is because Equation (25) can be expressed as a linear transform as follows:

$$C'(\mathbf{x},\mathbf{d}) = \frac{1}{|w_{\mathbf{k}}|}\sum_{\mathbf{x}\in w_{\mathbf{k}}}\left(a_{\mathbf{k}} I(\mathbf{x}) + b_{\mathbf{k}}\right), \quad (46)$$

where:

$$a_{\mathbf{k}} = (\Sigma_{\mathbf{k}} + \varepsilon U)^{-1}\left(\frac{1}{|w_{\mathbf{k}}|}\sum_{\mathbf{x}\in w_{\mathbf{k}}} I(\mathbf{x})C(\mathbf{x},\mathbf{d}) - \mu_{\mathbf{k}}\bar{C}(\mathbf{k},\mathbf{d})\right) \quad (47)$$

$$b_{\mathbf{k}} = \bar{C}(\mathbf{k},\mathbf{d}) - a_{\mathbf{k}}^{\mathrm{T}}\mu_{\mathbf{k}} \quad (48)$$

Here $\bar{C}(\mathbf{k},\mathbf{d})$ is the mean of the $\mathbf{d}-th$ slice of C within $w_{\mathbf{k}}$. Moreover, a factor that increases the speed of the guided image filter is that the summations in equations: Equation (46), Equation (47) and Equation (48) can be computed using box filters with a fixed window size [26]. Methodology B runs the guided image filtering for two fixed support windows sizes, therefore it can use box filters. On the other hand, the method in [27] that uses support windows of random sizes needs to estimate the summations for each pixel separately, an operation that increases the computational cost of [27].

The numeric disparity estimation results, after using in methodology B the scheme of [27] for performing guided image filtering, are given in the ``Adapt. Windows'' annotated column of Table 5. The differences between the column ``Best'' and ``Adapt. Windows'' prove that the proposed scheme for the definition of pixels support windows (see subsection 4.2.2.2) give better disparity evaluation results within methodology B.

6.6 Extended Comparison of both methodologies

Evaluation on just the four stereo pairs from the Middlebury online stereo database is not adequate to give a clear picture of the overall performance of an algorithm, since the average error rates of the best performing techniques are close to each other. Hence, the two proposed methodologies have been also evaluated on Dataset 2005 and Dataset 2006 (see Figure 33) in order to assess more extensively the performance of the proposed methodology. Dataset 2005 and Dataset 2006, which are presented in [71], include 27 stereo pairs with their ground truth.

Table 11 shows the results for the percentage of erroneous pixels having 1 or

2 disparity level difference with respect to ground truth. The results regarding the rest of methods in Table 11 are copied from the very recent work of [15]. The column ``All'' refers to case where all pixels on the disparity map are considered to estimate the percentage of erroneous pixels, while the term ``Visible'' refers to the case where only the pixels on the disparity map that correspond to unoccluded regions are considered to estimate the percentage of erroneous pixels. Methodology B gives slightly better results for the case of ``All'' regions and Δd>1, while it gives evidently better results for the case of ``Visible'' regions and Δd>1. However, Methodology A gives significantly better results than Methodology B for both ``All'' and ``Visible'' regions and Δd>2. This indicates that for methodology A the estimated disparity for some pixels is very close to their ground truth disparity and differs just 2 disparity levels.
In the Appendix the numeric error results for each of the 27 stereo pairs for both methodologies can be found. In particular, the numeric error results for methodology A and methodology B are given in Table 12 and Table 14, respectively. Additionally, the disparity maps for the 27 stereo pairs, with their respective disparity error maps for Δd>2, can be found for methodology A and methodology B in Table 13 and Table 15, respectively.

Error%	**Δd>1**	**Δd>1**	**Δd>2**	**Δd>2**
	All	**Visible**	**All**	**Visible**
Methodology A	12.13	8.26	7.64	4.74
Methodology B	12.07	7.71	8.32	5.07
Inf. Permeability[15]	14.15	7.98	10.34	6.46
Guided Filter[18]	15.06	8.40	11.82	6.80
Geodesic Support[30]	16.49	9.85	11.76	8.04
Var. Cross[23]	17.13	8.81	12.69	7.04
Adapt. sup.[21]	16.94	9.54	13.10	7.42

Table 11. The error results for the extended stereo datasets.

6.7 Summary

The current chapter presented the evaluation and the experiments related to methodologies A and B.

Chapter 6 – Evaluation and experiments

The computational analysis shows that methodology B is faster than methodology A and that for both methodologies the major portion of the computational cost is spent on the cost aggregation (C.A.) step. In order to evaluate the accuracy of methodologies A and B, the disparity maps generated for the Tsukuba, Venus, Teddy and Cones image pairs (after executing methodologies A and B using their optimum parameters) were submitted to the Middlebury online evaluation system. The ranking results indicate that methodology A and methodology B rank 4th and 5th, respectively, among 164 methods. Therefore, both methodologies have high disparity estimation accuracy. This fact is also confirmed from the evaluation of both methodologies on the 2005 and 2006 Middlebury Datasets, where methodologies A and B give more accurate results than several literature approaches tested on the same datasets.

Moreover, this chapter contains the numerical and the visual evaluations of the overall steps of methodologies A and B, which show how the quality of the disparity map gradually improves after applying sequentially methodologies' corresponding steps. Except for the overall evaluation of methodologies A and B, particular evaluations for several steps of these methodologies are also given.

Regarding methodology A, this chapter provides the numerical evaluation of the two phase combination strategy, which was introduced in the cost aggregation step of methodology A, and it also gives the numeric evaluation of the substeps that constitute the disparity optimization step. Further parameters testing for methodology A is included in the current chapter, too.

Regarding methodology B, this chapter includes experimental results showing that the exploitation of rectangular support windows of two sizes in the cost aggregation step helps to enhance the disparity estimation accuracy. Additionally, this chapter gives comparison results, which show that the cost aggregation approach proposed in methodology B outperforms a relevant cost aggregation method proposed in the literature. Further parameters testing for methodology B is provided in this chapter, too.

Chapter 7

7 Conclusions and Discussion

7.1 Conclusions

Methodology A produces very accurate disparity results for stereo image pairs. In order to achieve increased accuracy, this methodology uses efficiently three cost metrics to acquire a reliable combined cost volume. The optimization of the cost volume is performed using a semi-global optimization method, where a new criterion for the definition of the smoothness penalty terms is introduced, which helps to improve the disparity estimation results. Outliers handling is performed combining basic outlier handling and mean-shift segmentation based outlier handling. Another innovative aspect of this methodology is the way disparities are filtered based on histogram analysis in order to be used in uniform regions handling.

Methodology B, as well as methodology A, gives very accurate disparity results for stereo image pairs. This approach uses an efficient cost term, composed of three individual pixel-based cost terms, in order to estimate the initial cost volume. The filtered cost volume is acquired after applying image guided filtering to the initial cost volume, using rectangular support regions of two sizes. The optimization of the filtered cost volume is performed using weighted semi-global matching. Outliers handling is improved by introducing a straightforward scheme.

The high performance of the both methodologies method is verified experimentally using the Middlebury evaluation benchmark and an extended stereo dataset.

7.2 Discussion

This PhD proposes two methodologies that introduce novel ideas in the short-baseline stereo vision problem. In specific, methodology A, introduces an approach for acquiring a combined cost volume by exploiting three types of cost metrics. The first cost metric combines RGB-CENSUS information, the second one uses only CENSUS information and the third one SIFT information. The cost metrics are aggregated using

adaptive weights and their cost volumes are acquired. A reliable two-phase strategy is then followed to merge the individual cost volumes into a combined one. This approach, to the extent of my knowledge, is the first one that combines efficiently RGB, CENSUS and SIFT information in order to acquire a combined cost volume. Additionally, I did not come across any other method in this field that attempts to combine cost volumes that emerged from different cost metrics. Therefore, it is expected that this work could serve as guidance for other approaches, which will attempt to combine cost volumes that emerge from different cost metrics.

Additionally, methodology A introduces histogram analysis for handling uniform areas. This technique removes efficiently outlier disparities from large low-texture image regions, before applying disparity plane fitting in each region using the remaining reliable disparities. Furthermore, methodology A proposes an efficient strategy, which incorporates mean-shift segmentation-based outlier handling, to successfully cope with occluded areas.

Methodology B introduces a novel approach for exploiting guided image filtering to solve the stereo vision problem. In summary, the guided image filtering is applied separately for orthogonal support windows of two different sizes, where the appropriate support window size for each pixel is selected based on the texture homogeneity within the local region around this pixel. The texture homogeneity of a pixel is analyzed according to the mean-shift segment where the pixel belongs. Methodology B also exploits a simple, but efficient scheme, to successfully handle outliers. This scheme checks if the pixels on the right or on the left side of the outlier pixel are more similar in terms of colour to that pixel, before assigning a disparity value to it.

Both methodologies A and B have contributed to enhance the semi-global optimization technique. In specific, methodology A proposes the exploitation of mean-shift segmentation for introducing an additional criterion, which is used for the definition of the smoothness penalty terms. On the other hand, methodology B proposes a weighted variant of the semi-global optimization, where the path costs of a considered pixel may have different weights depending on the pixels that precede the considered pixel along each path direction.

Regarding the disparity estimation accuracy, both methodologies A and B rank

very high on the Middlebury online evaluation benchmark as it is evident in Table 6. In specific, methodology A ranks third among published methods, while methodology B ranks fourth among published methods. Table 11 also confirms the high accuracy of methodologies A and B. Methodology B gives slightly better accuracy for disparity level difference $\Delta d>1$ than methodology A, while methodology A gives significantly better accuracy for disparity level difference $\Delta d>2$ than methodology B. This fact shows that the disparity maps that are generated using methodology A have more pixels with disparity value that is close to their ground truth disparity value than the disparity maps that are generated using methodology B.

The computational cost of methodology B is significantly less than that of methodology A as it is deduced from Table 7. The computational burden of methodology A lies in the cost aggregation step which involves the computation of adaptive support weights. The other steps of methodology A are not computationally intensive.

7.3 Possible extensions and future work

Regarding methodology A, a future extension could be the adoption of a different cost aggregation approach, which will have low computational cost and at the same time will keep the disparity estimation accuracy at the same standards as the naive methodology A. Future work on both methodologies A and B could focus on improving the disparity estimation for very slanted surfaces, since it was not spent special effort on handling them. Future work could also involve the implementation of these methodologies in GPU, in order to improve their computational efficiency by taking advantage GPU's powerful parallel processing capabilities.

Section 2.6 presents some approaches that fuse depth sensors with stereo vision systems. As a future work, it would be interesting to examine the possibility to exploit particular elements of the proposed methodologies to develop such a fusion system.

Bibliography

[1] D. Scharstein and R. Szeliski, “A taxonomy and evaluation of dense two frame stereo correspondence algorithms”, International Journal of Computer Vision, vol. 47, no. 1-3, pp. 7-42, 2002.

[2] J. M. López-Valles, M.A. Fernández, A. Fernández-Caballero, M.T. López, J. Mira and A.E. Delgado, “Motion-based stereovision model with potential utility in robot navigation”, SPRINGER, Innovations in Applied Artificial Intelligence, Lecture Notes in Computer Science, vol. 3533, pp. 16-25, 2005.

[3] Z. Xiaozhou, L. Huimin, Y. Xingrui, L. Yubo and Z. Hui, “Stereo vision based traversable region detection for mobile robots using u-v-disparity”, CHINESE Control Conference, pp. 5785-5790, 2013.

[4] E. Gudis, G.V.D. Wal, S. Kuthirummal, S. Chai, S. Samarasekera, R. Kumar, and V. Branzoi, "Stereo Vision embedded system for Augmented Reality", In Conference on Computer Vision and Pattern Recognition Workshops, pp.15-20, 2012.

[5] M. Sizintsev, S. Kuthirummal, S. Samarasekera, R. Kumar, H.S. Sawhney and A. Chaudhry, “GPU Accelerated Realtime Stereo for Augmented Reality”, In International Symposium on Three-Dimensional Data Processing, Visualization, and Transmission, 2010.

[6] S. Gehrigm, “Large-field-of-view stereo for automotive applications”, In Proceedings of Workshop on Omnidirectional Vision, 2005.

[7] Z. Zhang, X. Ai, N. Canagarajah and N. Dahnoun, “Local stereo disparity estimation with novel cost aggregation for sub-pixel accuracy improvement in automotive applications”, IEEE Intelligent Vehicles Symposium, pp.99-104, 2012.

[8] E. Trucco and A. Verri, “Introductory Techniques for 3-D Computer Vision”, Prentice-Hall, 1998.

[9] S. Hermann and T. Vaudrey, “The gradient - A powerful and robust cost function for stereo matching”, International Conference of Image and Vision Computing New Zealand, pp. 1-8, 2010.

[10] S. Birchfield and C. Tomasi, “Depth discontinuities by pixel-to-pixel stereo”, International Journal of Computer Vision, vol. 35, pp. 269293, 1999.

[11] H. Liu, Y. Liu, S. OuYang, C. Liu and X. Li, “A novel method for stereo matching using gabor feature image and confidence mask”, In Visual Communications and Image Processing, pp. 16, 2013.

[12] W. Fife and J. Archibald, “Improved Census Transforms for Resource-Optimized Stereo Vision”, IEEE Transactions on Circuits and Systems for Video Technology, vol. 23, pp. 60-73, 2013.

[13] D. Min and K. Sohn, “An asymmetric post-processing for correspondence problem”, Elsevier, Signal Processing: Image Communication, vol. 25, no. 2, pp. 130-142, 2010.

[14] X. Sun, X. Mei, S. Jiao, M. Zhou and H. Wang, “Stereo matching with reliable disparity propagation”, In International Conference on 3D Imaging, Modeling, Processing, Visualization and Transmission, 2011.

[15] C. Cigla and A. A. Alatan, “Information permeability for stereo matching”, Elsevier Signal Processing: Image Communication, 2013.

[16] X. Mei, X. Sun, M. Zhou, S. Jiao, H. Wang and X. Zhang, “On building an accurate stereo matching system on graphics hardware”, In International Conference on Computer Vision Workshop on GPU in Computer Vision Applications, 2011.

[17] A. Klaus, M. Sormann, and K. Karner, “Segment-based stereo matching using belief propagation and a self-adapting dissimilarity measure”, In International Conference on Pattern Recognition, pp. 15-18, 2006.

[18] A. Hosni, C. Rhemann, M. Bleyer, C. Rother and M. Gelautz, “Fast Cost-Volume Filtering for Visual Correspondence and Beyond”, IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 35, no. 2, pp. 504-511, 2013.

[19] Z. Lee, J. Juang and T. Nguyen, “Local disparity estimation with three-moded cross census and advanced support weight”, IEEE Transactions on Multimedia, vol. 15, no. 8, pp. 1855-1864, 2013.

[20] F. Tombari, S. Mattoccia, L. Di Stefano and E. Addimanda, “Classification and Evaluation of Cost Aggregation Methods for Stereo Correspondence”, In IEEE Conference on Computer Vision and Pattern Recognition, pp. 1-8, 2008.

[21] K.-J. Yoon and I. S. Kweon, “Adaptive support-weight approach for correspondence search”, IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 28, pp. 650-656, 2006.

[22] L. Di Stefano, F. Tombari, S. Mattoccia, “Segmentation-Based Adaptive Support for Accurate Stereo Correspondence”, Proc. IEEE Pacific-Rim Symp. Image and Video Technology, 2007.

[23] K. Zhang, J. Lu and G. Lafruit, “Cross-based local stereo matching using orthogonal integral images”, IEEE Transactions on Circuits and Systems for Video Technology, vol. 19, no. 7, pp. 1073-1079, 2009.

[24] C. Tomasi and R. Manduchi, “Bilateral filtering for gray and color images”, In International Conference on Computer Vision, pp. 839-845, 1998.

[25] Q. Yang, “Recursive Bilateral Filtering”, In Proc. European Conference on Computer Vision, pp. 399-413, 2012.

[26] K. He, J. Sun and X. Tang, “Guided image filtering”, In Proc. European Conference on Computer Vision, pp. 1-14, 2010.

[27] Q. Yang, P. Ji, D. Li, S.J. Yao and M. Zhang, “Fast stereo matching using adaptive guided filtering”, Elsevier, Image and Vision Computing, vol. 32, no. 3, pp. 202-211, 2014.

[28] A. Hosni, M. Bleyer and M. Gelautz, “Secrets of adaptive support weight techniques for local stereo matching”, Elsevier, Computer Vision and Image Understanding, vol. 117, no. 6, pp. 620-632, 2013.

[29] Z. Ma, K. He, Y.Wei, J. Sun and Enhua Wu, “Constant Time Weighted Median Filtering for Stereo Matching and Beyond”, In International Conference on Computer Vision, pp. 49-56, 2013.

[30] A. Hosni, M. Bleyer, M. Gelautz, and C. Rhemann, “Local stereo matching using geodesic support weights”, In IEEE International Conference on Image Processing, pp. 2093-2096, 2009.

[31] V. Kolmogorov and R. Zabih, “Computing visual correspondence with occlusions using graph cuts”, In International Conference on Computer Vision, Vol. 2, pp. 508-515, 2001.

[32] M. Bleyer and M. Gelautz, “Graph-cut-based stereo matching using image segmentation with symmetrical treatment of occlusions”, Signal Processing: Image Communication, vol. 22, no. 2, pp. 127-143, 2007.

[33] Z. F. Wang and Z. G. Zheng, “A region based stereo matching algorithm using cooperative optimization”, In IEEE Conference on Computer Vision and Pattern Recognition, pp. 1-8, 2008.

[34] R. Brockers, “Cooperative stereo matching with color-based adaptive local support”, In Conference on Computer Analysis of Images and Patterns, pp. 1019-1027, 2009.

[35] L. Hong and G. Chen, “Segment-based stereo matching using graph cuts”, In IEEE Conference on Computer Vision and Pattern Recognition, pp. 74-81, 2004.

[36] Q. Yang, L. Wang, R. Yang, H. Stewenius and D. Nister, “Stereo matching with color-weighted correlation, hierarchical belief propagation and occlusion handling”, IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 31, pp. 492-504, 2009.

[37] H. Hirschmuller, “Accurate and efficient stereo processing by semi-global matching and mutual information”, In IEEE Conference on Computer Vision and Pattern Recognition, pp. 807-814, 2005.

[38] H. Hirschmuller, “Stereo processing by semiglobal matching and mutual information”, IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 30, pp. 328-341, 2008.

[39] S. Hermann and R. Klette, “Iterative semi-global matching for robust driver assistance systems”, In Asian Conference on Computer Vision, vol.3, pp. 465-478, 2012.

[40] S. Huq, A. Koschan and M. Abidi, “Occlusion filling in stereo: Theory and experiments”, Elsevier, Computer Vision and Image Understanding, vol. 117, no. 6, pp. 688-704, 2013.

[41] J. Zhu, L. Wang, J. Gao and R. Yang, “Spatial-temporal fusion for high accuracy depth maps using dynamic MRFs”, IEEE Transactions on Pattern Analysis and Machine Intelligence, vol.32, no. 5, pp. 899-909, 2010.

[42] Q. Yang, K.-H. Tan, W. B. Culbertson, and J. G. Apostolopoulos, “Fusion of active and passive sensors for fast 3d Capture”, IEEE International Workshop on Multimedia Signal Processing, 2010.

[43] G. Somanath, S. Cohen, B. Price and C. Kambhamettu, “Stereo+Kinect for High Resolution Stereo Correspondences”, International Conference on 3D Vision, 2013.

[44] S. Mattoccia, F. Tombari, and L. D. Stefano, “Stereo vision enabling precise border localization within a scanline optimization framework”, In Proc. Asian Conference on Computer Vision, pp. 517-527, 2007.

[45] M. Humenberger, C. Zinner, M. Weber, W. Kubinger and M. Vincze, “A fast stereo matching algorithm suitable for embedded real-time systems”, Elsevier, Computer Vision and Image Understanding, vol. 114, no. 11, pp. 1180-1202, 2010.

[46] D. Comanicu and P. Meer, “Mean shift: A robust approach toward feature space analysis”, IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 24, pp. 603-619, 2002.

[47] N. Y. Chang, T. Tsai, B. Hsu, Y. Chen and T. Chang, “Algorithm and Architecture of Disparity Estimation With Mini-Census Adaptive Support Weight”, IEEE Transactions on Circuits and Systems for Video Technology, vol. 20, no. 6, pp. 792-805, 2010.

[48] H. Bay, T. Tuytelaars and L. Van Gool, “SURF: Speeded Up Robust Features”, In Proceedings of the European Conference on Computer Vision, 2006.

[49] Y. S. Heo, K. M. Lee and S. U. Lee, “Joint Depth Map and Color Consistency Estimation for Stereo Images with Different Illuminations and Cameras”, IEEE Transactions on Pattern Analysis and Machine Intelligence, vo. 35, no. 5, pp. 1094-1106, 2013.

[50] G. Saygili, L.J.P. van der Maaten and E.A. Hendriks, “Feature-Based Stereo Matching Using Graph Cuts”, In ASCI, 2011.

[51] L. Xu, O. C. Au, W. Sun, Y. Li, J. Li, “Hybrid Plane Fitting for Depth Estimation”, In Asia-Pacific Signal and Information Processing Association Annual Summit and Conference, 2012.

[52] M. Fischler and R. Bolles, “Random Sample Consesus: A Paradigm for Model Fitting with Applications to Image Analysis and Automated Cartography”, Communications of the ACM, vol. 6, pp. 381-395, 1981.

[53] J. Kowalczuk, E. Psota and L. C. Prez, “Real-Time Stereo Matching on CUDA Using an Iterative Refinement Method for Adaptive Support-Weight Correspondences”, IEEE Transactions on Circuits and Systems for Video Technology, vol. 23, no. 1, pp. 94-104, 2013.

[54] L. De-Maeztu, S. Mattoccia, A. Villanueva and R. Cabeza, “Efficient aggregation via iterative block-based adapting support-weights”, In International Conference on 3D Imaging, 2011.

[55] I. Jung, T. Chung, J. Sim and C. Kim, “Consistent Stereo Matching Under Varying Radiometric Conditions”, IEEE Transactions on Multimedia, vol. 15, no. 1, pp.56-69, 2013.

[56] S. Kumar, C. Micheloni, C. Piciarelli and G.L. Foresti, “Stereo rectification of uncalibrated and heterogeneous images”, Elsevier Pattern Recognition Letters, vo.31, no. 11, pp. 1445-1452, 2010.

[57] S. Kim and M. Pollefeys, “Robust radiometric calibration and vignetting correction”, IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 30, no. 4, pp. 562–576, 2008.

[58] S. Kagarlitsky, Y. Moses and Y. Hel-Or, “Piecewise-consistent color mappings of images acquired under various conditions”, In IEEE International conference on computer vision, pp. 2311-2318, 2009.

[59] Y. S. Heo, K. M. Lee and S. U. Lee, “Joint Depth Map and Color Consistency Estimation for Stereo Images with Different Illuminations and Cameras”, IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 35, no. 5, pp. 1094-1106, 2013.

[60] Edge Detection and Image Segmentation (EDISON) System: http://coewww.rutgers.edu/riul/research/code/EDISON/doc/ref.html.

[61] T. Liu, P. Zhang and L. Luo, “Dense stereo correspondence with contrast context histogram, segmentation-based two-pass aggregation and occlusion handling”, Proc. IEEE Pacific-Rim Symp. Image and Video Technology, 2009.

[62] P. Meer and B. Georgescu, “Edge detection with embedded confidence”, IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 23, pp. 1351-1365, 2001.

[63] C. Christoudias, B. Georgescu and P. Meer, “Synergism in low-level vision”, In International Conference on Pattern Recognition, vol. 4, pp. 150-155, 2002.

[64] J.R.R. Uijlings, A.W.M. Smeulders and R.J.H. Scha, “Real-Time Visual Concept Classification”, IEEE Transactions on Multimedia, vol 12, pp. 665-682, 2010.

[65] J. M. Geusebroek, A. W. M. Smeulders, and J. van de Weijer," Fast anisotropic gauss filtering", IEEE Transactions on Image Processing, vol. 12, no. 8, pp. 938-943, 2003.

[66] M. Gong, R.G. Yang, W. Liang, and M.W. Gong, "A performance study on different cost aggregation approaches used in real-time stereo matching", International Journal of Computer Vision, vol. 75, pp. 283-296, 2007.

[67] X. Hu and P. Mordohai, "Evaluation of stereo confidence indoors and outdoors", In IEEE Conference on Computer Vision and Pattern Recognition, pp. 1466-1473, 2010.

[68] Middlebury Stereo Evaluation: http://vision.middlebury.edu/stereo/.

[69] S. Mattoccia, S. Giardino and A. Gambini, "Accurate and efficient cost aggregation strategy for stereo correspondence based on approximated joint bilateral filtering", In Asian Conference on Computer Vision, pp. 371-382, 2009.

[70] M. Hubert, P.J. Rousseeuw and K. Vanden Branden, "ROBPCA: a new approach to robust principal components analysis", Technometrics, vol. 47, pp. 64-79, 2005.

[71] H. Hirschmuller and D. Scharstein, "Evaluation of cost functions for stereo matching", In IEEE Conference on Computer Vision and Pattern Recognition, pp. 1-8, 2007.

[72] K. Zhang, G. Lafruit, R. Lauwereins and L. Gool, "Joint integral histograms and its application in stereo matching", In International Conference on Image Processing, pp. 817-820, 2010.

[73] C. Banz, H. Blume and P. Pirsch, "Real-time semi-global matching disparity estimation on the GPU", In International Conference on Computer Vision Workshops, pp.514-521, 2011.

[74] Y. Zhan, Y. Gu, K. Huang, C. Zhang, and K. Hu, "Accurate image-guided stereo matching with efficient matching cost and disparity refinement. Submitted to IEEE Transactions on Circuits and Systems for Video Technology.

[75] M. Mozerov and J. van Weijer, "Accurate stereo matching by two step global optimization", submitted to IEEE Transactions on Image Processing, 2014.

[76] J. Liu, C. Li, F. Mei, and Z. Wang, “3D entity-based stereo matching with ground control points and joint second order smoothness prior”, SPRINGER, The Visual Computer, 2014.

[77] Y. Peng, G. Li, R. Wang, and W. Wang, “Stereo matching with space-constrained cost aggregation and segmentation-based disparity refinement”, Proc. SPIE 9393, Three-Dimensional Image Processing, Measurement , and Applications 2015.

[78] Y. Furukawa and J. Ponce, “Accurate, Dense and Robust Multi-View Stereopsis”, IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 32, pp. 1362-1376, 2010.

[79] V. Hiep, R. Keriven, P. Labatut and J. Pons, “Towards high-resolution largescale multi-view stereo”, In IEEE Conference on Computer Vision and Pattern Recognition, pp. 1430-1437, 2009.

[80] E. Tola, C. Strecha and P. Fua, “Efficient large-scale multi-view stereo for ultra high-resolution image sets”, Machine Vision and Applications, vol. 32, pp. 903-920, 2012.

[81] N. Salman and M. Yvinec, “Surface Reconstruction from Multi-View Stereo of Large-Scale Outdoor Scenes”, International Journal of Virtual Reality, 2010.

[82] N. Snavely, S. Seitz and R. Szeliski, “Modeling the world from internet photo collections”, International Journal of Computer Vision, vol. 80, pp. 189-210, 2008.

[83] A. Fusiello and L.Irsara, “Quasi-Euclidean epipolar rectification of uncalibrated images”, Machine Vision and Applications, vol. 22, pp. 663-670, 2010.

[84] E. Tola, V. Lepetit and P. Fua, “Daisy: an efcient dense descriptor applied to wide baseline stereo”, IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 32, pp. 815-830, 2010.

[85] M. Muja and D. Lowe, “Fast Approximate Nearest Neighbors with Automatic Algorithm Configuration”, International Conference on Computer Vision Theory and Applications, 2009.

[86] M. Levin, “Mesh-independent surface interpolation”, Geometric Modeling for Scientific Visualization, Springer-Verlag, pp. 37-49, 2003.

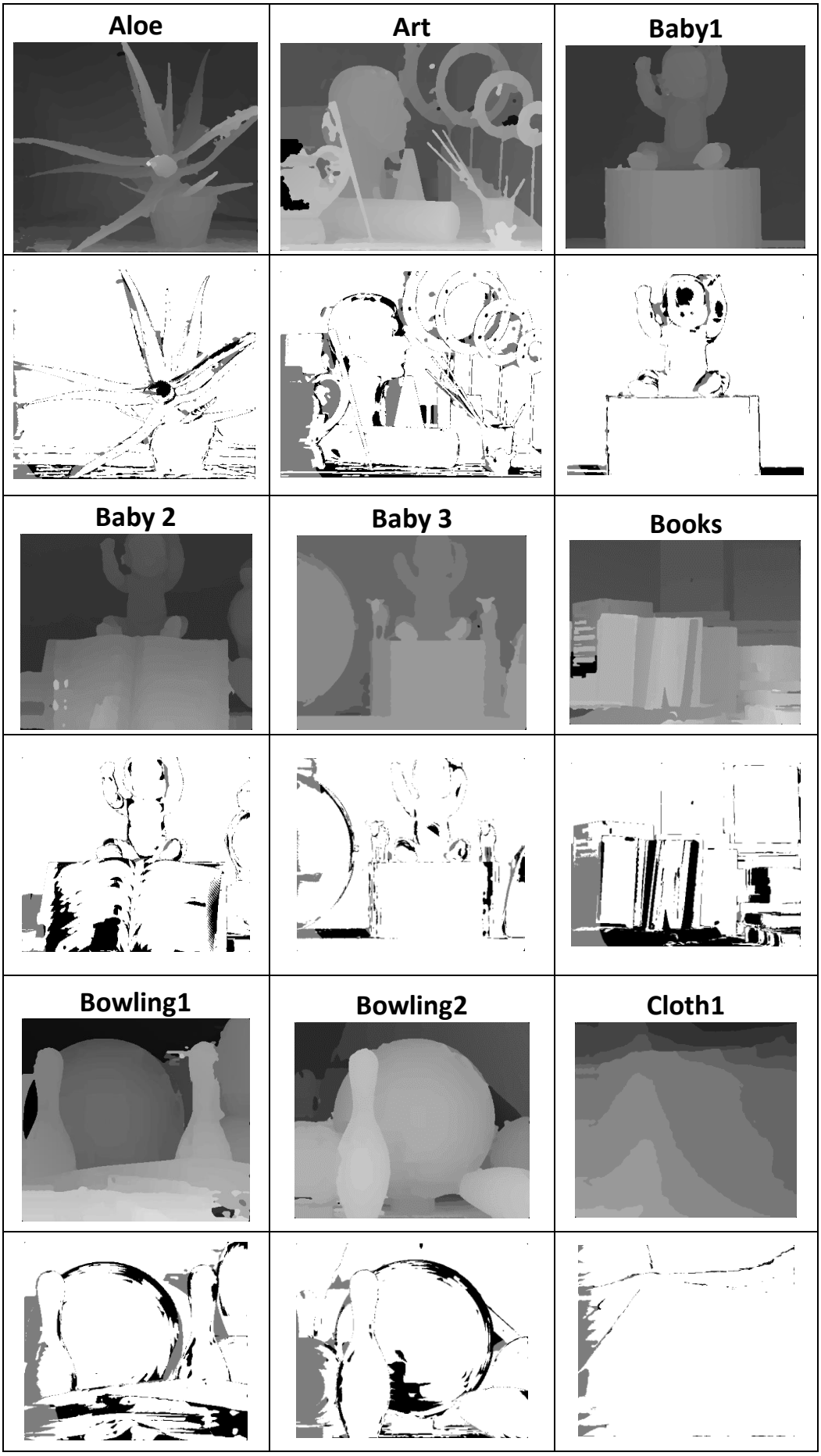
[87] C. Strecha, W. von Hansen , L. Van Gool, P. Fua, U. Thoennessen, “On Benchmarking camera calibration and multi-view stereo for high resolution imagery”, In IEEE Conference on Computer Vision and Pattern Recognition, 2008.

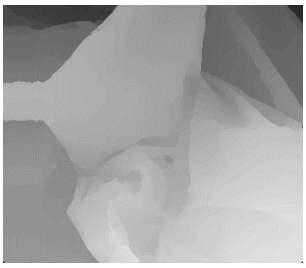
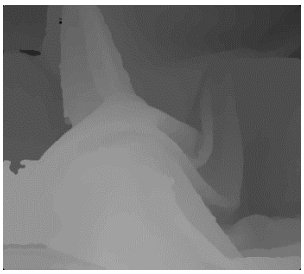
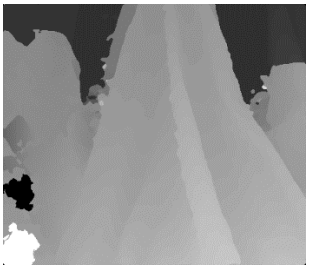
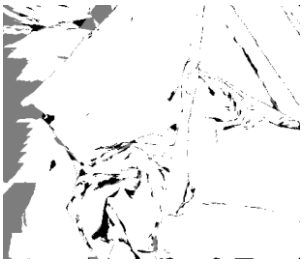
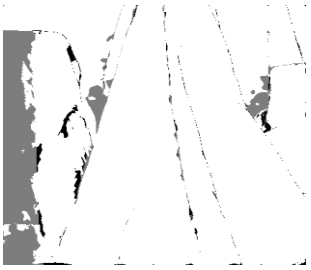
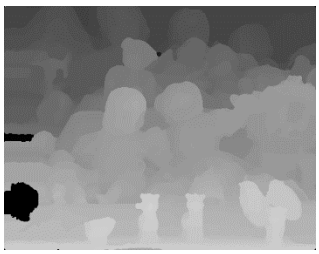
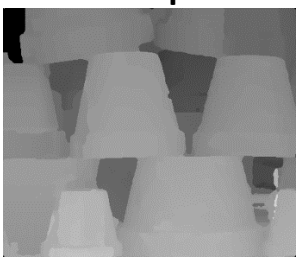
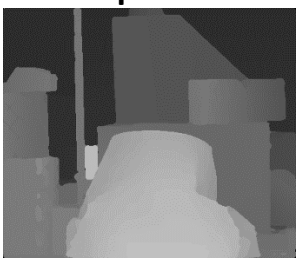
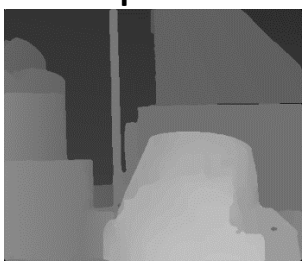
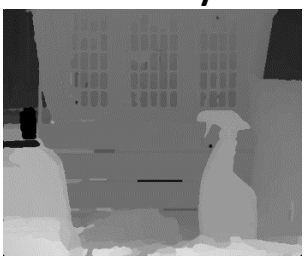
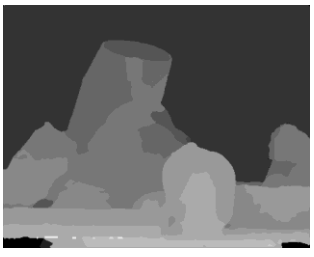
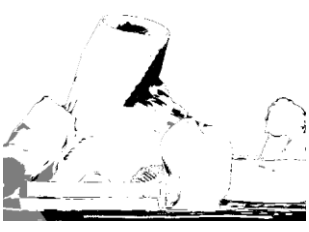
Appendix A – Results on the extended dataset

Results on the extended dataset for methodology A

Error%	Δd>1	Δd>1	Δd>2	Δd>2
	All	Visible	All	Visible
Aloe	7.97	4.903	5.49	3.17
Art	18.23	8.516	13.655	4.62
Baby1	5.33	4.158	3.198	1.948
Baby2	11.704	10.218	6.191	4.278
Baby3	8.26	5.793	4.579	2.577
Books	22.769	17.222	17.47	12.885
Bowling1	20.334	13.469	12.602	7.147
Bowling2	17.128	10.635	9.753	4.98
Cloth1	4.489	0.625	1.899	0.311
Cloth2	10.532	3.467	6.635	1.619
Cloth3	4.5467	1.518	3.16	0.822
Cloth4	11.986	1.721	9.005	0.695
Dolls	13.85	7.836	8.091	3.344
Flowerpots	26.81	23.348	16.529	14.531
Lampshade1	10.86	6.84	4.586	3.541
Lampshade2	11.212	8.763	6.24	5.623
Laundry	18.44	11.798	10.594	5.972
Midd1	9.69	7.46	6.291	4.697
Midd2	6.709	4.966	4.281	3.308
Moebius	13.207	8.97	8.985	5.526
Monopoly	16.918	16.72	15.881	15.872
Plastic	28.337	31.827	12.76	14.287
Reindeer	7.375	3.961	4.288	2.016
Rocks1	6.949	2.597	4.172	1.06
Rocks2	5.605	1.735	3.264	0.66
Wood1	4.557	3.439	3.062	2.148
Wood2	3.801	0.409	3.507	0.3575
Average	12.13	8.26	7.64	4.74

Table 12. Analytical error results for the extended stereo datasets using methodology A.


Aloe
Art
Baby1
Baby 2
Baby 3
Books
Bowling1
Bowling2
Cloth1

Cloth2	Cloth3	Cloth4
		
		
Dolls	Flowerpots	Lampshade1
		
		
Lampshade2	Laundry	Midd1
		
		

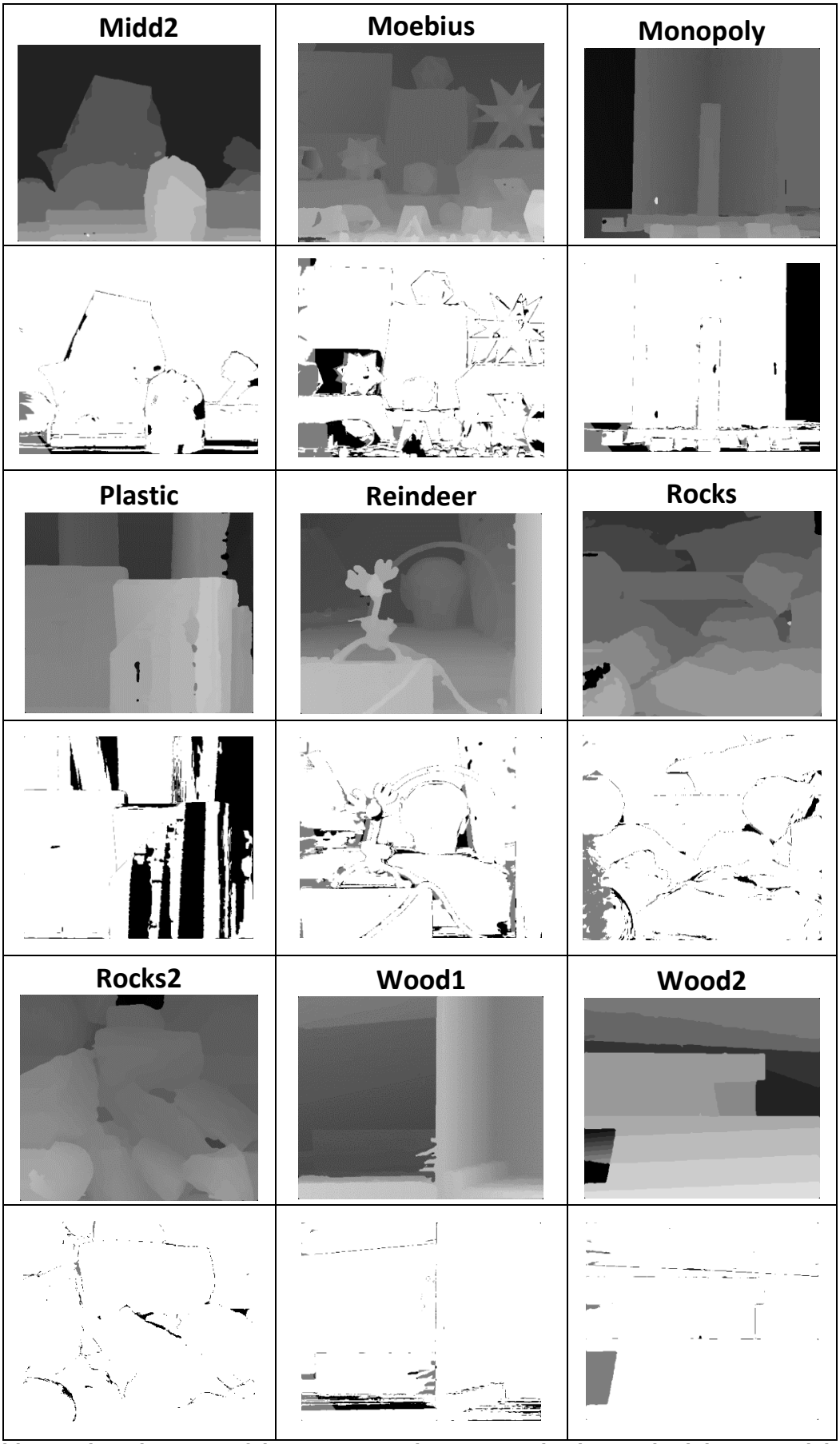
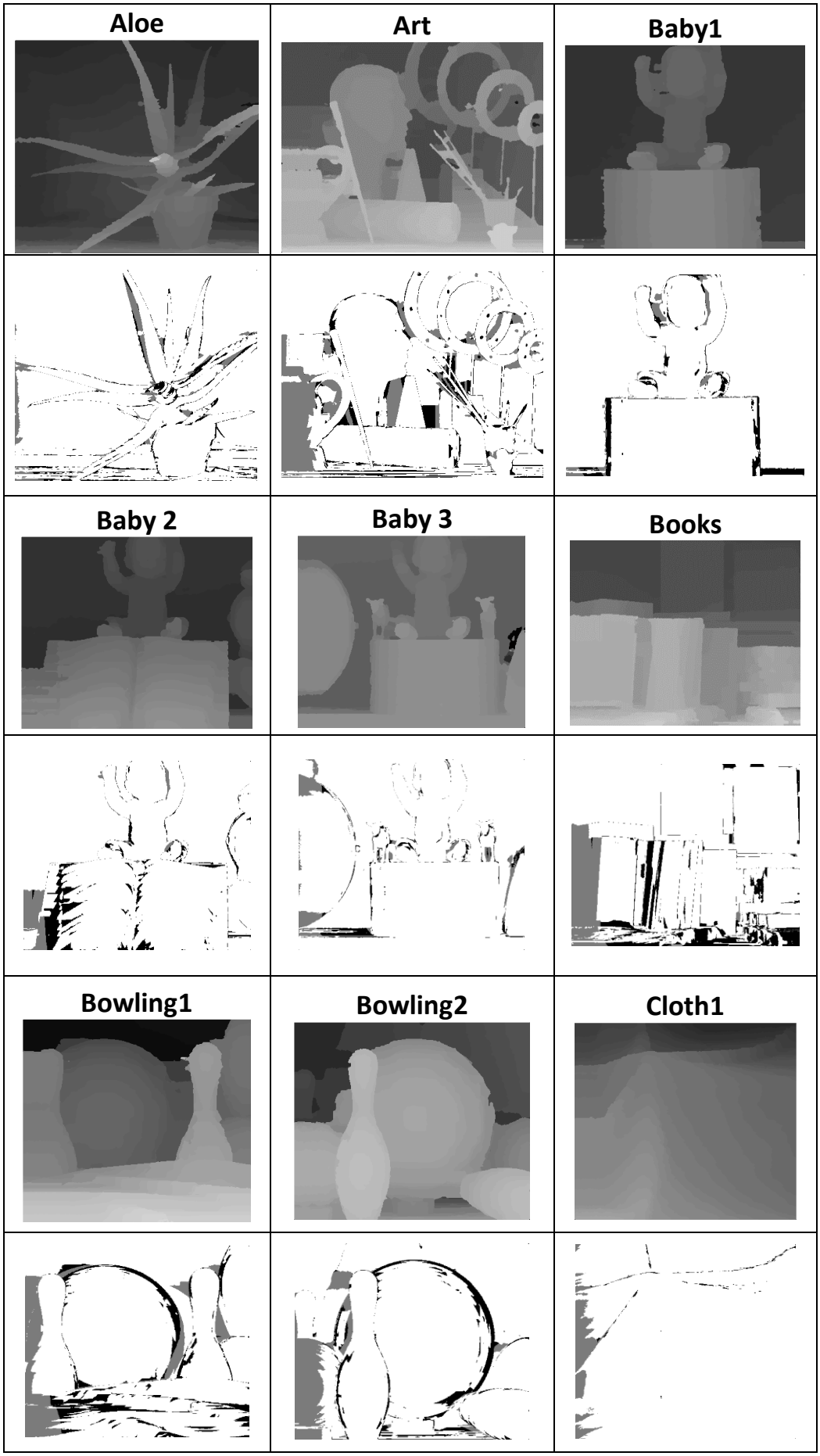


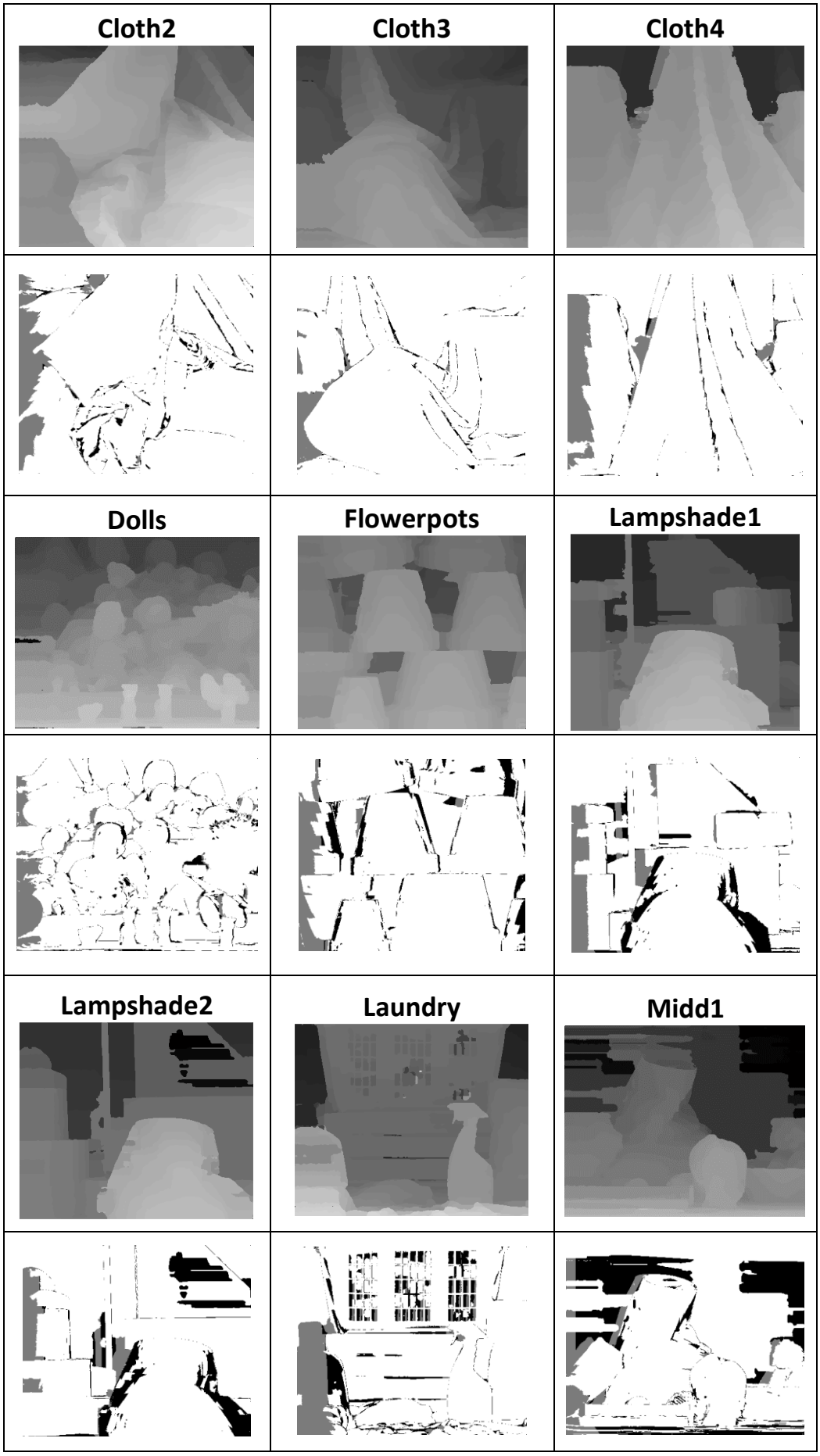
Table 13. Disparity maps of the 27 stereo pairs generated using methodology A and their corresponding disparity error maps for error threshold 1.

Results on the extended dataset for methodology B

Error%	Δd>1	Δd>1	Δd>2	Δd>2
	All	Visible	All	Visible
Aloe	7.002	4.139	4.688	2.59
Art	16.975	7.364	13.56	4.827
Baby1	5.218	3.982	3.414	2.196
Baby2	7.842	5.316	4.954	2.277
Baby3	6.065	2.92	4.522	1.986
Books	17.41	10.536	12.395	5.588
Bowling1	17.168	9.217	10.532	3.79
Bowling2	14.073	6.475	8.259	2.61
Cloth1	4.298	0.387	2.02	0.24
Cloth2	10.685	3.177	7.06	1.48
Cloth3	4.955	1.65	3.48	0.846
Cloth4	10.63	1.621	6.656	0.928
Dolls	10.80	4.497	6.946	2.205
Flowerpots	15.616	9.145	7.58	3.56
Lampshade1	14.825	8.418	8.926	6.09
Lampshade2	17.042	12.924	12.482	10.82
Laundry	19.64	11.104	15.073	7.27
Midd1	26.164	22.256	22.63	19.43
Midd2	19.88	16.125	16.504	13.80
Moebius	12.51	8.526	9.23	5.98
Monopoly	12.41	10.873	10.04	9.187
Plastic	30.628	34.172	20.95	23.32
Reindeer	8.063	5.345	4.726	2.92
Rocks1	4.802	1.849	2.563	0.766
Rocks2	4.60	1.296	2.706	0.51
Wood1	5.78	4.573	2.448	1.465
Wood2	0.724	0.262	0.267	0.258
Average	12.07	7.71	8.32	5.07

Table 14. Analytical error results for the extended stereo datasets using methodology B.


Aloe
Art
Baby1
Baby 2
Baby 3
Books
Bowling1
Bowling2
Cloth1


Cloth2
Cloth3
Cloth4
Dolls
Flowerpots
Lampshade1
Lampshade2
Laundry
Midd1

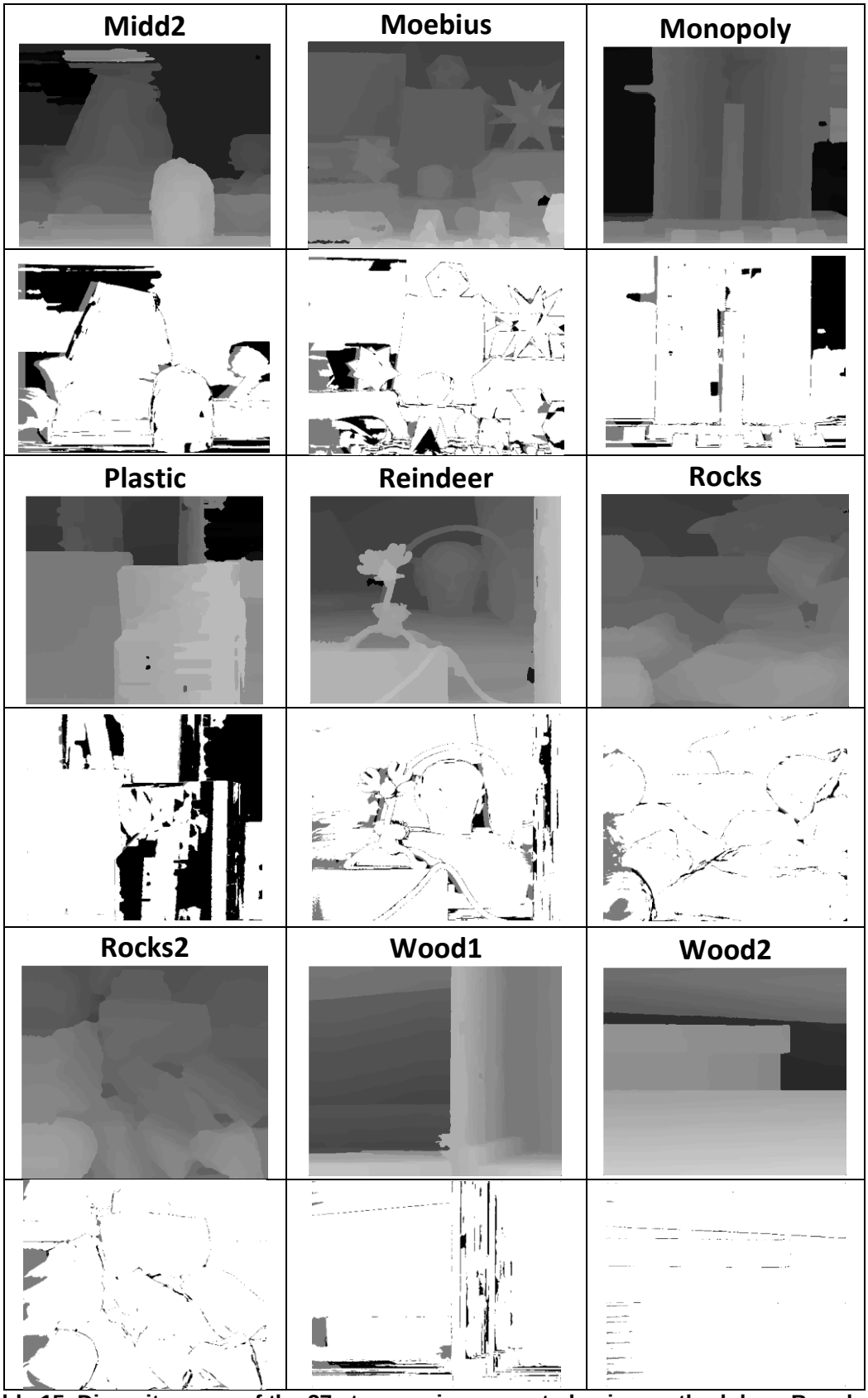


Table 15. Disparity maps of the 27 stereo pairs generated using methodology B and their corresponding disparity error maps for error threshold 1.

Appendix B – Case Study: Wide-baseline stereo matching and point cloud generation

The main objective of this PhD thesis is to present solutions for addressing the short-baseline stereo vision problem. Thus, the largest part of this thesis is devoted to describe relevant methodologies.

However, during the PhD period, some research effort was devoted to develop an approach aiming to increase the accuracy of the stereo point clouds, which are generated from stereo pairs with wide-baseline. Appendix B presents this approach.

B.1 Introduction

The automatic and accurate 3D modeling of objects and scenes, from multiple photographs or videos, constitutes an important objective in the computer vision and graphics research fields. The realistic 3D models can be exploited in multiple applications, such as computer graphics, TV/film special effects and computer games.

Research in 3D model reconstruction, using multi-view stereo algorithms, has made significant progress in the computer vision community. Multi-view stereo (MVS) algorithms take multiple images with pose information as input and produce dense 3D models with increased accuracy.

Several of the algorithms generate and merge collections of 3D points clouds, which may be then used to generate a mesh surface [78], [79], [80], [81]. Many of the algorithms that rely on 3D point clouds, put emphasis on the merging of the point clouds that are generated from different stereo pairs by using visibility constraints to filter erroneous points.

The approach presented in this chapter could foster these algorithms by improving the accuracy of the individual point clouds, which are generated from each wide-baseline stereo pair, before point clouds from all stereo pairs are merged.

The proposed approach is described in section B.2. While, section B.3 provides information on the parameters used, the experimental results and the computational cost.

B.2 Stereo dense 3D point cloud generation

In general, the first step of multi-view 3D reconstruction is the computation of camera(s) poses that capture a scene. The Structure-from-Motion (SfM) approach presented in [82] provides an efficient way for computing robustly the camera parameters from a set of user-generated images.

In this appendix, an efficient methodology for generating an accurate 3D point cloud from a stereo image pair, is presented. The approach can be divided into three stages:

- During the first stage, the stereo pairs to be used for the generation of each stereo point cloud are appropriately selected, based on specific conditions, in order to ensure the accuracy of reconstruction.
- The second stage includes the estimation of dense correspondences between the images of the stereo pair, based on DAISY [84] descriptor matching. Additionally, a strategy for filtering outlier correspondences is presented.
- The third stage involves the refinement of the generated 3D point cloud. Refinement is accomplished by estimating the correspondences in sub-pixel accuracy and by smoothing the resulting point cloud using the moving least squares algorithm.

The innovation of this methodology lies mainly in the efficient strategy for removing outliers and in the effective combination of sub-pixel accuracy correspondences estimation with the moving least squares algorithm to improve the accuracy of the generated 3D point cloud. In the following, more details are provided on what each of these stages comprises.

B.2.1 Stereo pair selection

Stereo images pair selection is a crucial step to acquire stereo 3D point clouds with good accuracy. The images of an ``adequate'' stereo pair should have significant overlap to be easily matched, but also to be sufficiently separated, since much closeness may result to point cloud estimation errors. This is quantified, similarly to

[80], by measuring the angle θ between the camera principal rays of the stereo images. The condition that θ should satisfy is: $\theta_{min} < \theta < \theta_{max}$.

Afterwards, Quasi-Euclidean epipolar rectification [83] is applied to each stereo images pair that satisfies the previous condition. If the Quasi-Euclidean epipolar rectification error T_{rect} is below a threshold T_{max}, the stereo pair is assumed as suitable for proceeding to the estimation of its point cloud. Consequently, this work, except for the condition based on θ, defines a second condition based on T_{rect} for selecting adequate stereo pairs.

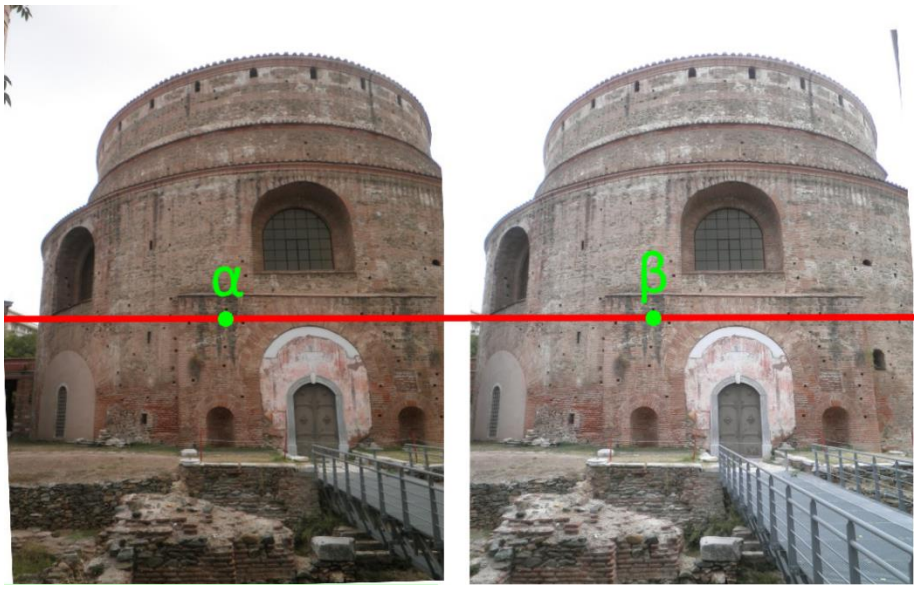


Figure 40. Correspondence in rectified stereo images.

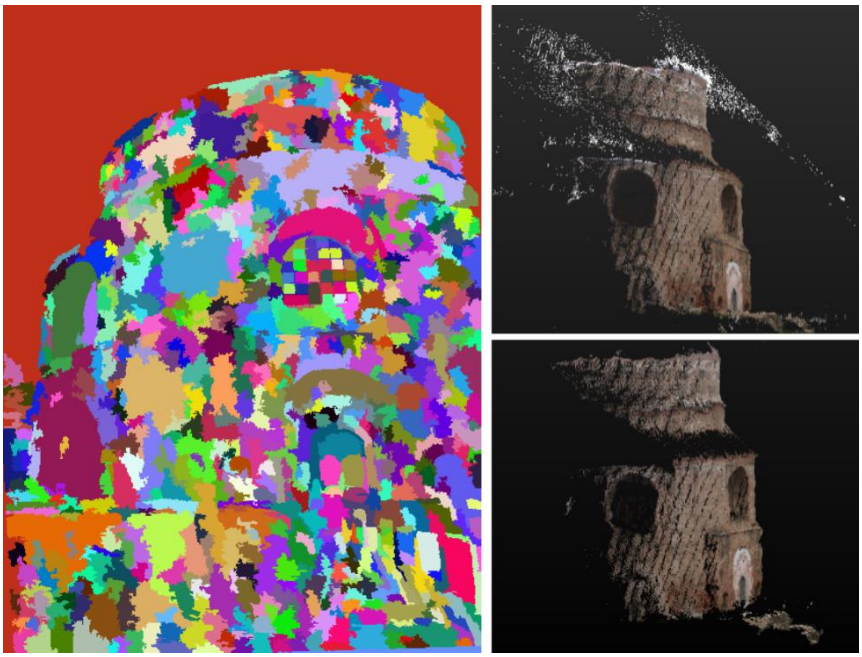
B.2.2 Dense correspondences estimation and outliers filtering

During the second step, the DAISY descriptor [84] is exploited to estimate dense correspondences between the images of a stereo pair. DAISY has been selected for this scope, because it has been proved in [84] to be very efficient for dense wide baseline matching.

More specifically, in order to find for a pixel on one image its corresponding pixel to the other image, for the pixel's DAISY descriptor the pixel with the nearest DAISY descriptor on the second image is searched. The search is constrained along horizontal epipolar lines, since the images have been rectified using Quasi-Euclidean epipolar rectification [83]. Figure 40 depicts a pixel correspondence α-β on an

epipolar line, between a rectified stereo pair. The search for the nearest descriptor is performed using approximate nearest neighbor searching based on randomized kd-trees [85], where trees are searched in parallel. The kd-trees search approach significantly boosts the speed of searching, when compared to exhaustive search.

The correspondence estimation is performed twice. Once having as reference the first image of the stereo pair and once having as reference the second image. Then, the Left-Right consistency check [44] is used for detecting the correspondence outliers.



(a) (b)

Figure 41. Illustration of: (a) left image's mean-shift segmentation map, (b) the generated stereo point cloud without using (upper part) and, when using (bottom part) the proposed outliers filtering strategy.

Except for this common technique, an additional technique for filtering outliers in a segment level, and not in a pixel level, is proposed. This technique helps to remove outliers that appear in low-textured regions. Initially, mean-shift segmentation is used to partition the image into different segments that contain groups of pixels (the segmentation map of the left image of Figure 40 is visualized in Figure 41a).

Then for each segment, the percent of pixels that pass the right-left consistency check to the total number of pixels contained in the segment, is computed. If this percent is over 50%, then the correspondences in the segment are considered as inliers. Otherwise, all the correspondences in the segment are considered as

outliers.

This strategy assists in filtering numerous outliers. This fact is evident in the visual example of Figure 41b. The upper part of Figure 41b shows the point cloud that is generated without using the proposed outliers filtering strategy, while the bottom part of Figure 41b depicts the point cloud after applying the outliers filtering strategy. Obviously, the second point cloud contains less outliers.

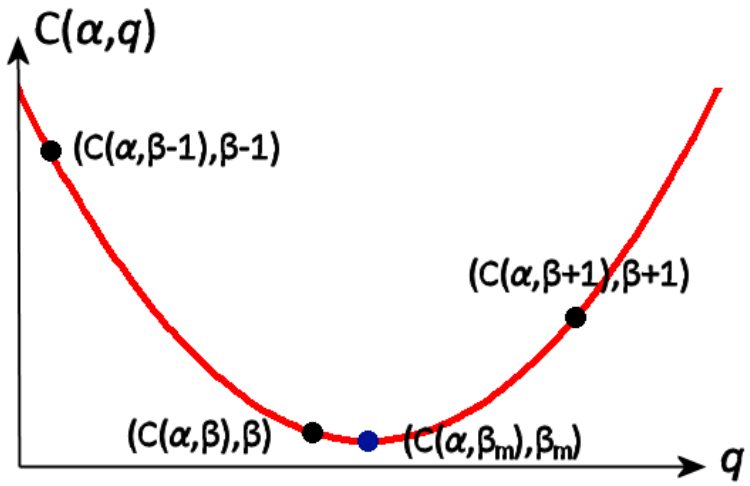


Figure 42. Sub-pixel accuracy correspondence using quadratic curve fitting.

B.2.3 Point cloud refinement

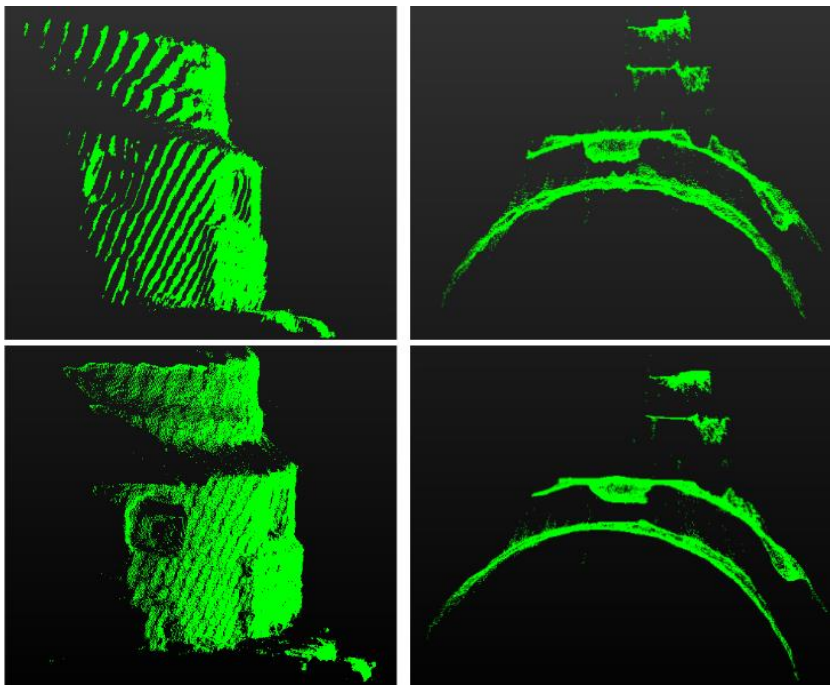
B.2.3.1 Correspondences estimation in sub-pixel accuracy

So far, the estimated correspondences have pixel accuracy. However, correspondence estimation at sub-pixel accuracy can significantly improve the quality of the generated 3D point cloud, since pixel accuracy matching, results in discrete and not continuous values of depth information.

In order to achieve sub-pixel accuracy the following process is followed. Let us suppose that a pixel α on the left image corresponds to a pixel β on the right image and their matching cost $C(\alpha,\beta)$ has already been estimated. Then, the matching cost $C(\alpha,\beta\text{-}1)$ between the DAISY descriptors of pixels α and β-1 and the matching cost $C(\alpha,\beta+1)$ between α and $\beta+1$ are estimated. The three points $(C(\alpha,\beta\text{-}1),\beta\text{-}1)$, $(C(\alpha,\beta),\beta)$ and $(C(\alpha,\beta+1),\beta+1)$ (these points are visualized in Figure 42) are used to

estimate a quadratic function and estimate the minimum cost $C(\alpha,\beta_m)$ of the quadratic function's curve, which corresponds to β_m . Consequently, the sub-pixel accuracy correspondence is assumed to be given by the pair (α,β_m) , while the pixel accuracy correspondence was given by the pair (α,β) .

The upper part of Figure 43a shows the point cloud that corresponds to pixel accuracy correspondences, while the bottom part of Figure 43a depicts the point cloud that corresponds to sub-pixel accuracy correspondences. It is evident, by comparing these two parts, that the bottom point cloud is more accurate, since depth information is continuous.



(a) (b)

Figure 43 Illustration of: (a) the point cloud that corresponds to pixel accuracy (upper part) and sub-pixel accuracy (bottom part) correspondences, (b) the point cloud before (upper part) and after (bottom part) applying the Moving Least Squares algorithm.

B.2.3.2 Point cloud smoothing

The estimated 2D sub-pixel correspondences are converted into 3D point clouds using the projection matrices that were estimated during the SfM process. Afterwards, a final step is applied to improve the reconstruction quality.

More specifically, in order to resample and smooth the generated point cloud the Moving Least Squares (MLS) algorithm, described in [86], is exploited. The upper

part of Figure 43b shows the point cloud before applying the MLS algorithm, while the bottom part of Figure 43b after applying the MLS algorithm. Evidently, the bottom point cloud is more accurate.

B.3 Experimental results

B.3.1 Set of optimum parameters

The limits for the principal rays' angles are set to $\theta_{min} = 5^{\circ}$ and $\theta_{max} = 25^{\circ}$. The rectification error threshold is set to $T_{max} = 0.5 \cdot (D_{max} / 640)$ pixels, where D_{max} is the maximum dimension of the images (width or height), which constitute the image pair, in pixels. In this way, T_{max} is set proportional to the size of the stereo images to be rectified.

The selected parameters for computing the DAISY descriptor are the radius of the descriptor $R = 9$, the number of rings $Q = 3$, the number of histograms on each ring $T = 4$ and the number of bins of the histograms $H = 4$.

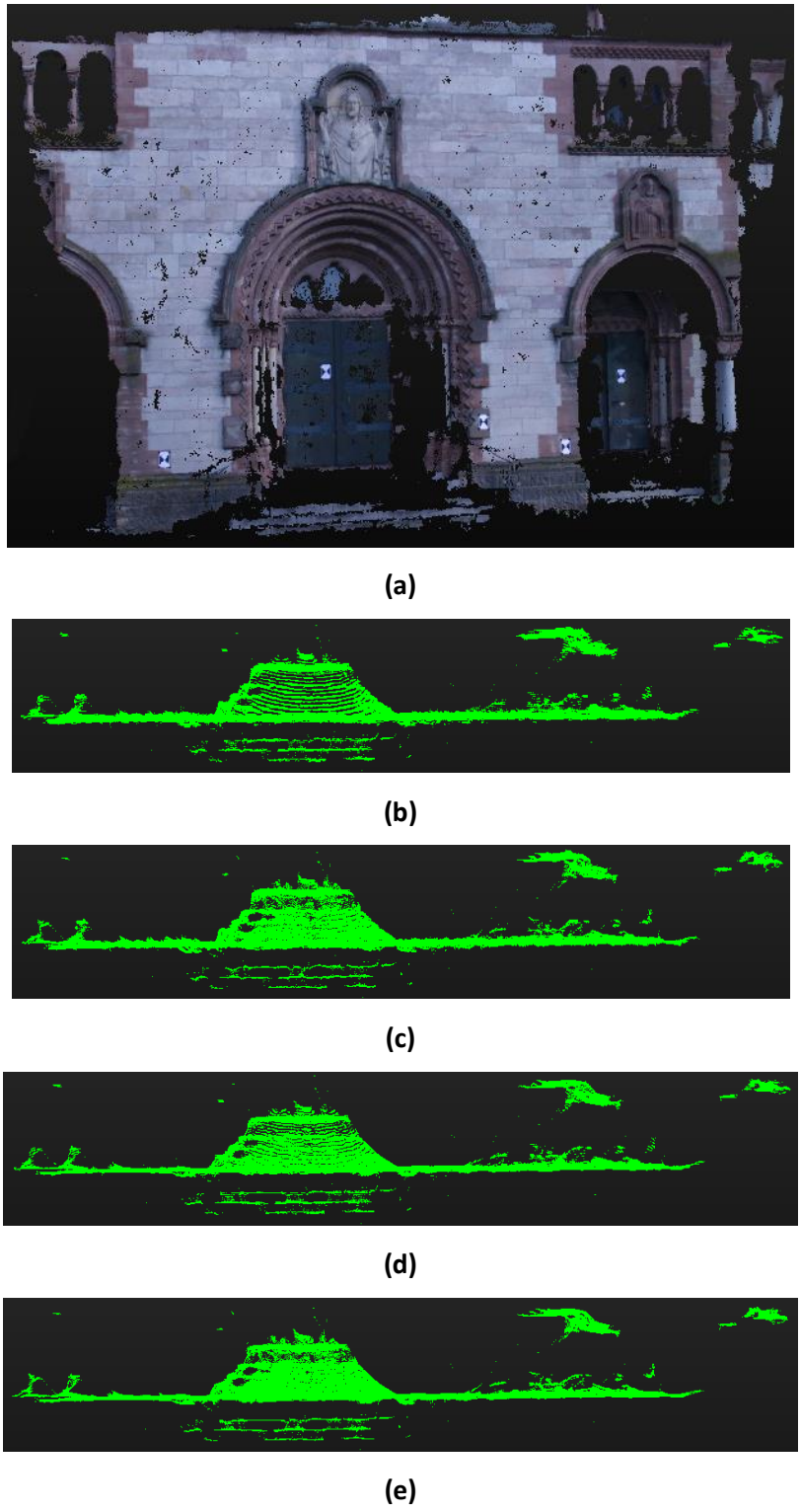
The parameters used for the mean-shift segmentation are the segmentation spatial radius σ_s, which is set to σ_s = 3 and the segmentation feature space radius σ_r, which is set to σ_r = 3.

B.3.2 Experiments

A stereo pair of images, which has been derived from the Herz-Jesu-P8 [87] (in specific images ``0007.png'' and ``0008.png'') is used to visually indicate the improvement introduced by the proposed methodology, regarding the accuracy of the estimated stereo point cloud. The images have been downscaled with a factor of 3, so as to make more obvious the accuracy improvement in visual data of lower resolution. The generated stereo 3D point cloud using this approach is visualized in Figure 44a.

In the following, the stereo point cloud, with or without using the proposed refinement steps, is estimated. The point cloud (observed from the upper viewpoint): (i) without using sub-pixel accuracy nor MLS algorithm is visualized in Figure 44b, (ii) using only sub-pixel accuracy is visualized in Figure 44c, (iii) using only MLS algorithm is visualized in Figure 44d and (iv) using both sub-pixel accuracy and MLS algorithm is

visualized in Figure 44e. Evidently, Figure 44e gives the more accurate stereo point cloud.



(a)

(b)

(c)

(d)

(e)

Figure 44. Illustration of (a) the colored stereo point cloud. The generated point cloud using: (b) neither sub-pixel accuracy nor MLS, (c) only sub-pixel accuracy, (d) only MLS, (e) sub-pixel accuracy and MLS.

Figure 45. Rotunda 3D Reconstruction

In the second example, the proposed methodology is used to generate individual stereo point clouds using images captured from the Rotunda Ancient Monument in the city of Thessaloniki. Then, the point clouds are finally concatenated to form the final 3D point cloud. This 3D reconstruction example is depicted in Figure 45. The right part of Figure 45, which depicts the overview of the Rotunda 3D reconstruction, indicates that individual point clouds have satisfactory accuracy, so that they are well registered to form a complete 3D representation of the captured object, even without using any method for combining the individual point clouds.

B.4 Discussion and future work

The methodology, presented in this case study, which assists in generating accurate point clouds from wide-baseline stereo pairs, could be exploited by multi-view algorithms, which attach great importance to the combination of sets of stereo point clouds and not to the computation of the individual stereo point clouds. In specific, these algorithms could be fostered by using the proposed methodology in order to improve the accuracy of the individual stereo point clouds, before point clouds from all stereo pairs are merged. For instance, the method in [80] uses a complex methodology that verifies the accuracy of each 3D point on more multiple depth maps and does not give weight to the individual stereo point cloud computation.

Future work could examine the exploitation of the presented methodology within a general framework that will also contain an approach for the efficient combination of individual stereo point clouds.

B.5 Summary

This case study presents a time efficient and accurate methodology for generating 3D point clouds of good accuracy from wide-baseline stereo pairs. Initially, the methodology defines some conditions for the proper selection of image pairs. Then, the selected stereo images are used to estimate dense correspondences using the Daisy descriptor. An efficient two-phase strategy to remove outliers is then introduced. Finally, the 3D point cloud is refined by combining sub-pixel accuracy correspondences estimation and the moving least squares algorithm. The experimental results show that this methodology assists in acquiring point clouds of better accuracy when compared to the point clouds that are generated using descriptor-based matching in pixel accuracy.

Publications

International Journals

- G. Kordelas, D. Alexiadis, P. Daras, E. Izquierdo, "Enhanced disparity estimation in stereo images", ELSEVIER, Image and Vision Computing, accepted for publication.
- S. Essid, X. Lin, M. Gowing, G. Kordelas, A. Aksay, P. Kelly, T. Fillon, Q. Zhang, A. Dielmann, V. Kitanovski, R. Tournemenne, A. Masurelle, E. Izquierdo, N. E. O'Connor, P. Daras, G. Richard, " A multi-modal dance corpus for research into interaction between humans in virtual environments ", Journal on Multimodal User Interfaces, Special Issue on Multimodal Corpora, Springer, 2012.

International Conferences Proceedings

- G. Kordelas, D. Alexiadis, P. Daras, E. Izquierdo, "Revisiting guided image filter based stereo matching and scanline optimization for improved disparity estimation", IEEE International Conference on Image Processing, 2014.
- G. Kordelas, P. Daras, P. Klavdianos, E. Izquierdo, Q. Zhang, "Accurate stereo 3D point cloud generation suitable for multi-view stereo reconstruction", IEEE International Conference on Visual Communications and Image Processing, 2014.
- D. Alexiadis, G. Kordelas, K. Apostolakis, J. Agapito, J. Vegas, E. Izquierdo, P. Daras, "Reconstruction for 3D Immersive Virtual Worlds", Workshop on Image Analysis for Multimedia Interactive Services, 2012.

International Journals Under Review

- G. Kordelas, D. Alexiadis, P. Daras, E. Izquierdo, “Content-based guided image filtering and weighted semi-global optimization for fast and accurate disparity estimation”, submitted to IEEE Transactions on Multimedia.