UNIVERSIDAD NACIONAL DE COLOMBIA

Método de selección automática de algoritmos de correspondencia estéreo en ausencia de *ground truth*

Camilo José Vargas Cortés

Universidad Nacional de Colombia

Facultad de Minas, Departamento de Ciencias de la Computación y de la Decisión

Medellín, Colombia

2015

Método de selección automática de algoritmos de correspondencia estéreo en ausencia de *ground truth*

Camilo José Vargas Cortés

Tesis de maestría presentada como requisito parcial para optar al título de:

Maestría en Ingeniería - Ingeniería de Sistemas

Director (a):
John Willian Branch Bedoya
Doctor en ingeniería de Sistemas
Codirector (a):
Iván Mauricio Cabezas Troyano
Doctor en ingeniería con énfasis en computación

Línea de Investigación:
Visión por computador
Grupo de investigación:
GIDIA

Universidad Nacional de Colombia
Facultad de Minas, Departamento de Ciencias de la Computación y de la Decisión
Medellín, Colombia
2015

Resumen

La correspondencia estéreo es un campo ampliamente estudiado que ha recibido una atención notable en las últimas tres décadas. Es posible encontrar en la literatura un número considerable de propuestas para resolver el problema de correspondencia estéreo. En contraste, las propuestas para evaluar cuantitativamente la calidad de los mapas de disparidad obtenidos a partir de los algoritmos de correspondencia estéreo son relativamente escasas. La selección de un algoritmo de correspondencia estéreo y sus respectivos parámetros para un caso de aplicación particular es un problema no trivial dada la dependencia entre la calidad de la estimación de un mapa de disparidad y el contenido de la escena de interés.

Este trabajo de investigación propone una estrategia de selección de algoritmos de correspondencia estéreo a partir de los mapas de disparidad estimados, por medio de un proceso de evaluación en ausencia de *ground truth*. El método propuesto permitiría a un sistema de visión estéreo adaptarse a posibles cambios en las escenas al ser aplicados a problemas en el mundo real. Esta investigación es de interés para investigadores o ingenieros aplicando visión estéreo en campos de aplicación como la industria.

Palabras clave: visión estéreo, algoritmos de correspondencia estéreo, estimación de mapas de disparidad, metodologías de evaluación, selección de parámetros.

Abstract

The stereo correspondence problem has received significant attention in literature during approximately three decades. A plethora of stereo correspondence algorithms can be found in literature. In contrast, the amount of methods to objectively and quantitatively evaluate the accuracy of disparity maps estimated from stereo correspondence algorithms is relatively low. The application of stereo correspondence algorithms on real world applications is not a trivial problem, mainly due to the existing dependence between the estimated disparity map quality, the algorithms parameter definition and the contents on the assessed scene.

In this research a stereo correspondence algorithms selection method is proposed by assessing the quality of estimated disparity maps in absence of *ground truth*. The proposed method could be used in a stereo vision to increase the system robustness by adapting it to possible changes in real world applications. The contribution of this work is relevant to researchers and engineers applying stereo vision in fields such as industry.

Keywords: stereo vision, stereo correspondence algorithms, disparity map estimation, evaluation methods, parameter selection.

Contents

Pág.

Introduction

The stereo vision or stereopsis is the process of estimating three-dimensional information from a stereo image pair. The 3D depth information is estimated based on the difference of horizontal coordinates of corresponding pixels on the stereo image pair. This process is performed naturally by the human system, which translates the stereo images into a three-dimensional perception of the scene (Scharstein, 1999). The stereo vision problem is an inverse and ill-posed for two main reasons: the structural ambiguity on input images, caused by repetitive patterns; and the ambiguity introduced by similarity measures.

Stereo correspondence algorithms take as input a rectified image pair, and compute a disparity map as output. The estimation of an accurate disparity map still remains a challenging task, mainly due to the presence of occluded pixels, and textureless regions, among other factors inherent to the problem (Z.-F. Wang & Zheng, 2008).

Finding the corresponding points for both images is the problem addressed by stereo correspondence algorithms. Difficulties solving this problem includes: matching ambiguities, due to repetitive patterns or locally uniform intensities; occlusion, when only a single projection of a 3D point is captured into the stereo image pair; the assumption of equal intensity values for corresponding points, which means that the scene is composed of Lambertian surfaces and there are no camera bias or gain differences; among others (I. Cabezas, 2013; Scharstein, 1999).

Stereo correspondence has several application fields, such as autonomous navigation (Morales & Klette, 2009), pedestrian detection (Keller, Enzweiler, & Gavrila, 2011) and agriculture (Nielsen, Andersen, Slaughter, & Granum, 2007). The most noticeable difficulties when using stereo correspondence algorithms on real world applications is the fact that the quality of estimated disparity maps depends of the content of the evaluated scene. In the same way, the definition of appropriate input parameters of the stereo correspondence algorithms also depends on the contents of the scene.

A plethora of stereo correspondence algorithms can be found in literature, where different approaches are proposed. In contrast, the amount of methods to objectively and quantitatively evaluate the accuracy of disparity maps estimated from stereo correspondence algorithms is relatively low. This situation becomes more evident for evaluation approaches where no ground truth information is available. Nevertheless, an assessment on the progress of stereo correspondence can only be achieved if quantitative and objective performance results are reported for proposed algorithms (Scharstein & Szeliski, 2002).

Stereo correspondence evaluation methods are classified as ground-truth based methods or methods performed in the absence of ground-truth (I. Cabezas, 2013). Ground-truth based methods rely on independent measurements by active sensors (Scharstein et al., 2014; Scharstein & Szeliski, 2003). Disparity maps resulting from stereo correspondence algorithms are compared against ground-truth information using metrics such as Bad Matched Pixels (BMP)(Scharstein & Szeliski, 2002), Bad Matched Pixels Relative Errors (BMPRE) (I. Cabezas, Padilla, & Trujillo, 2012), SZE (I. Cabezas, Padilla, & Trujillo, 2011), among others. Evaluation methods in the absence of ground-truth estimate disparity maps quality by computing errors on predicted views (Szeliski, 1999) or using confidence metrics (Haeusler & Klette, 2012). In practice, many researchers on the area might be relying blindly on a single evaluation method, ignoring which their strengths and weaknesses really are.

Synthetic data has been used in quantitative evaluation due to the difficulties to generate ground-truth on real imagery. However, synthetic data may fail to model the complexities of real-world, or in contrary, be artificially of a high complexity (Scharstein & Szeliski, 2003). In fact, the generation of disparity ground-truth may be too difficult or laborious and even impossible to achieve in some circumstances due to the limitations of active stereo techniques to be used in indoor or controlled environments (Morales & Klette, 2011).

Evaluation methods that do not use ground truth data can be classified as prediction error approaches and confidence measure approaches (Morales & Klette, 2011; B.-S. Shin, Caudillo, & Klette, 2015). The prediction error approach (Szeliski, 1999) suggest the prediction of a novel view of the scene. The predicted view can be compared to a reference view obtained from a third camera in a known position. However, error scores reflect not only the accuracy of the disparity estimation algorithm, but also the accuracy of the selected rendering algorithm, since the rendering process of the predicted view has to deal with interpolation or extrapolation issues (Scharstein & Szeliski, 2002; Sellent & Wingbermühle, 2012). Confidence metrics are used to measure the reliability of the estimated disparity value for each pixel (Morales & Klette, 2011). Several stereo correspondence algorithms use confidence metrics as part of their estimation processes in order to refine the resulting disparity maps.

This work is motivated by the above mentioned issues. Here, the results of a research oriented to select adequate stereo correspondence algorithms by assessing the quality of their output disparity maps in absence of ground truth are presented. The contribution of this work is relevant to researchers and engineers applying stereo vision in fields such as the industry. The proposed approach presented in this thesis allows the objective selection of a stereo correspondence algorithm and its respective parameters to estimate the disparity map of a static scene. Here, the higher quality disparity map estimated from a fixed set of stereo correspondence algorithms and their respective parameters can be selected for a near real-time application, where the contents of the assessed scene could change in the process. The proposed approach is compared against standard ground truth evaluation methods in an online framework.

1. Problem definition

In the stereo vision context is possible to acquire a rectified stereo image pair $(\mathrm{I_l}, \mathrm{I_r})$ of a specific scene E. Afterwards, a disparity map MD can be estimated by using any stereo correspondence algorithm F and its respective parameters P.

A plethora of stereo correspondence algorithms can be found in literature. The Middlebury stereo evaluation website alone has more than 160 stereo correspondence algorithms available. The selection of an appropriate algorithm in absence of *ground truth* to reconstruct a scene in a specific application field is a non-trivial problem, in particular, when the scene contents change in time.

A quality metric Q is required in order to perform the selection of a stereo correspondence algorithm by assessing the estimated disparity maps $\overline{MD}$. This quality metric is computed for each disparity map as shown in equation 1:

$$\forall\, MD_i \in \overline{MD}, \exists\, !\, Q_i \in \overline{Q} \mid Q_i = g\left(I_{l_i}, I_{r_i}, md\right) \tag{1}$$

Where, g is a function used to compute each Q_i in absence of ground truth and will be defined as part of this thesis. The set of computed quality metrics $\bar{Q}$ allows the selection of a stereo correspondence algorithm as shown in equation 2:

$$\max_{Q_i \in \bar{Q}} (Q) \tag{2}$$

Figure 1.1 shows a simplified scheme of the relevant elements in the problem definition. Here, a set of disparity maps is estimated for the scene using a fix set of stereo correspondence algorithms and its respective parameters. Each disparity map is assessed in absence of *ground truth* computing a quality metric Q. Finally, the quality metric allows the selection of a stereo correspondence algorithm with presumably the lower error rate.

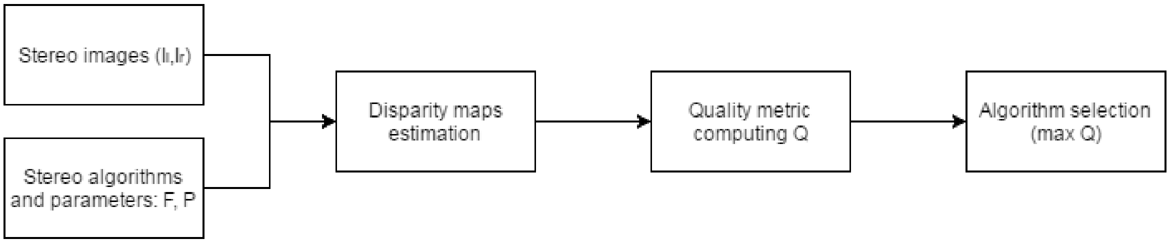

Stereo images (I_l,I_r)
Stereo algorithms and parameters: F, P
Disparity maps estimation
Quality metric computing Q
Algorithm selection (max Q)

Fig 1.1. Scheme of relevant elements in the problem definition.

The research presented in this document is based on the following research question: How a stereo correspondence algorithm with its respective parameters can be chosen to produce a disparity map for a given scene with presumably the lower error rate, where no ground truth information or additional views are available?

2. Theoretical background

2.1 Image formation

In this work the pinhole camera model will be used to explain the image formation process, since this camera model resembles closely the operation of modern cameras. The main difference between the pinhole camera model and modern cameras is that modern cameras use lenses to focus light into an array of sensors for image acquisition (Scharstein, 1999). The pinhole camera model is composed of a box or dark chamber with a small hole in one side. The light of the observed scene pass through the pinhole creating a reversed image on the back of the box. Figure 2.2.1 shows an illustration of a pinhole camera.

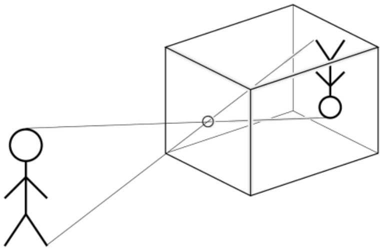


Fig. 2.1.1. Pinhole camera illustration.

Here, the pinhole is the optical center denoted by c. The front of the box will represent the focal plane. And the image plane I, will be located at the back side of the box, at a distance f from the focal plane. The relation between 3D scene coordinates and 2D image coordinates can be established using perspective projection and homogeneous coordinates (also called perspective coordinates). The 2D image coordinate system (x, y) is defined at the optical center c of the pinhole camera. The 3D scene coordinate system

(X,Y,Z) is defined at the center of the image plane I. Figure 2.1.2 shows the coordinate system for the pinhole camera model. The Z axes for both coordinate systems coincide. The X,Y axes on the scene coordinate system are parallel to x,y axes on the image coordinate system respectively.

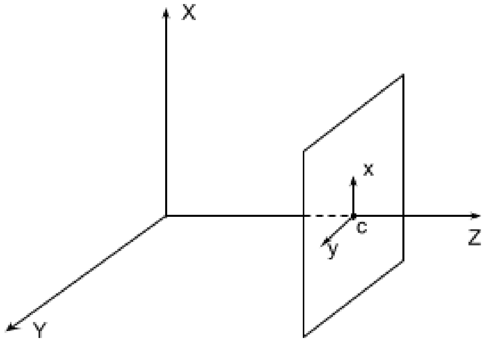


Fig. 2.1.2. Pinhole camera model.

The relation between 3D scene and 2D image coordinate systems is given by:

$$\frac{x}{X} = \frac{y}{Y} = \frac{f}{Z} \qquad (3)$$

Hence, an arbitrary 3D point in the scene coordinate system can be expressed in the image coordinate system as:

$$x = \frac{Xf}{Z}, \qquad y = \frac{Yf}{Z} \qquad (4)$$

In modern cameras the optical, analog image transformation into a digital image, is done using a rectangular grid of sensors, where the intensity distribution on the image plane is quantized into integer values. This yields the known image representation as a 2D array of discrete values of intensity called pixels (Scharstein, 1999).

2.2 Stereo vision

Stereo vision or stereopsis is the process of estimating 3D information of a scene using two slightly different 2D images acquired simultaneously. The depth information of a point in the 3D space is estimated based on the change in position of this point between the two images. This process is performed without effort by the human system, which translates the stereo images into a three-dimensional perception of the scene [1]. In the following sections the concepts of stereo correspondence are explained.

2.2.1 Stereo geometry

The geometry used to represent a stereo acquisition system is called epipolar geometry. The epipolar geometry presents the relationship between a physical point and its projection into the left and right image planes of the stereo system as shown in figure 2.2.1.1.

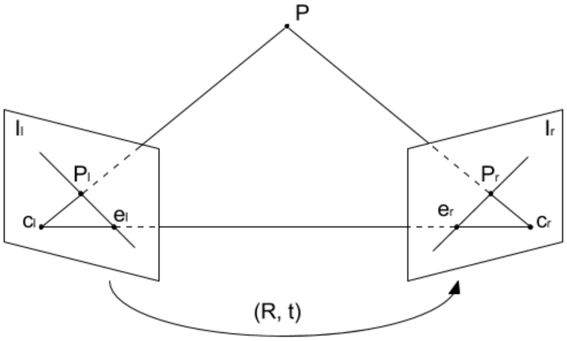


Fig. 2.2.1.1. Epipolar geometry

Where, c_l and c_r are the optical centers of the left and right view respectively. P is a physical point in the 3D scene and P_l, P_r its respective projections on the left and right image planes I_l, I_r. e_l and e_r called the epipoles are defined by the intersection of the line defined by c_l , c_r and the image planes I_l, I_r. The lines defined by the epipoles and the projections of P are called the left and right epipolar lines. Here, each 2D image plane captures the projection of the physical point P from the 3D scene as explained in the pinhole camera model. Given the projection P_l of a physical point P on the image plane I_l, its corresponding projection P_r in the right image plane I_r will lie along the corresponding epipolar line. This is known as the epipolar constraint. This reduces the correspondence search from 2D to 1D, which is considerably useful in establishing correspondences. Figure 2.2.1.2 shows a search representation for a projection P_l on the image I_r using the epipolar constraint.

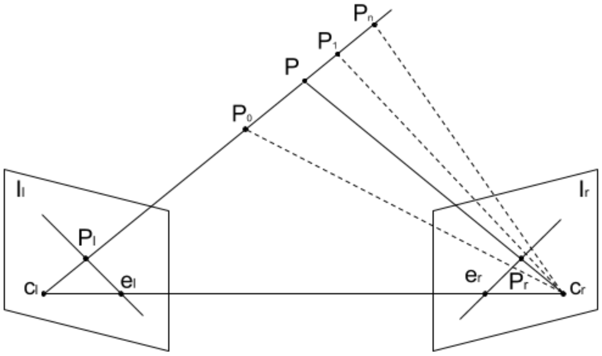


Fig. 2.2.1.2. Epipolar constraint

The epipolar constraint means that the possible two-dimensional search for matching features across two images becomes a one-dimensional search along the epipolar lines. This is not only a vast computational savings; it also allows us to reject a lot of points that could otherwise lead to spurious correspondences (Bradski & Kaehler, 2008).

The epipolar geometry for a pair of cameras is implicit in the relative pose and calibrations of the cameras, and can easily be computed from point matches using the fundamental matrix (Szeliski, 2010). Once this geometry has been computed, we can use the epipolar line corresponding to a pixel in one image to constrain the search for corresponding pixels in the other image. A more efficient and simple approach is to adjust the stereo acquisition system and apply rectification on the input images so that epipolar lines are horizontal. Here, the geometry computation can be ignored and the search is performed in the x axis. Figure 2.2.1.3 shows a simplified stereo system.

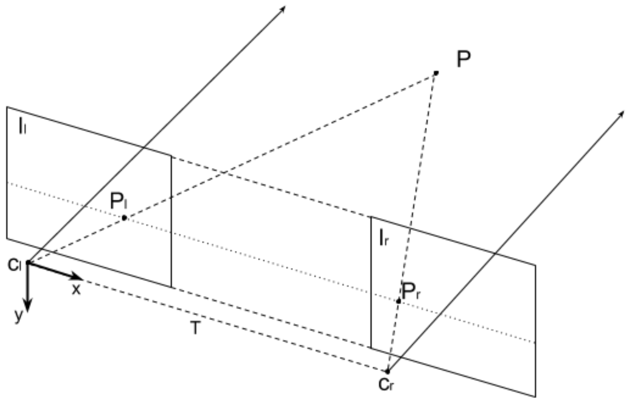


Fig. 2.2.1.3. Stereo system simplified to parallel image planes and horizontal epipolar line.

In this approach the optical axes are parallel and perpendicular to the image planes. In addition the epipolar lines are parallel to the x axis. In practice, a perfectly aligned configuration is rare within a real stereo system; hence, a rectification process must be performed. This rectification process is done via image warping, using estimations of intrinsic and extrinsic camera parameters. Further information about the rectification process is presented in (Bradski & Kaehler, 2008; Szeliski, 2010).

2.2.2 Stereo correspondence problem

The stereo correspondence problem is defined as the estimation of 3D information of a scene from a pair of 2D images. This estimation is performed based on the distance of corresponding points on the input pair stereo images. Figure 2.2.2.1 presents a basic stereo system, used to explain the use of corresponding points in depth estimation.

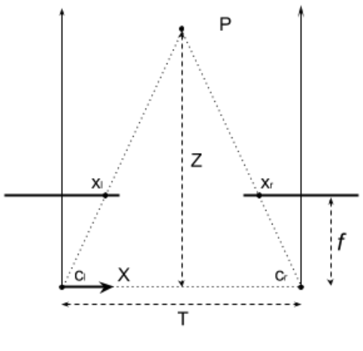


Fig. 2.2.2.1. Stereo correspondence.

Where P is a physical point projected into left and right images to x_l and x_r coordinates. c_l and c_r are the optical centers of the left and right views. f is the focal distance for the stereo cameras. T is the distance between the optical centers and Z is the depth of the point P to the stereo system baseline. The difference between the left and right x coordinates is defined as disparity as shown in equation 5.

$$d = x_l - x_r \qquad (5)$$

In this simplified case, using similar triangles allow to demonstrate that depth Z is inversely proportional to the disparity between the views as follows:

$$\frac{T-d}{Z-f} = \frac{T}{Z}, \qquad d = \frac{fT}{Z} \qquad (6)$$

In practice, the information of corresponding points coordinates is unknown. Finding the corresponding points for both images is the problem addressed by stereo correspondence algorithms. Difficulties solving this problem includes: matching ambiguities, due to repetitive patterns or locally uniform intensities; occlusion, when only a single projection of a 3D point is captured into the stereo image pair; the assumption of equal intensity values for corresponding points, this means that the scene is composed of Lambertian surfaces and there are no camera bias or gain differences; among others (I. Cabezas, 2013; Szeliski, 2010).

The stereo correspondence problem is an ill-posed problem due to the lack of information about depth and the instability of the solution of the system. As a consequence of instability, a small perturbation in the matching of conjugated points may produce large errors in the 3D information recovery process (I. Cabezas, 2013).

Solving the stereo correspondence problem for a few feature points in the stereo image pair is called non-dense (or sparse) stereo correspondence. This work focuses in dense stereo correspondence algorithms and evaluation approaches, where the quantity of matched pixels between images is expected to be high in relation to the images resolution. The section 2.2.3 introduces the methods used to solve the stereo correspondence problem.

2.2.3 Stereo correspondence approaches

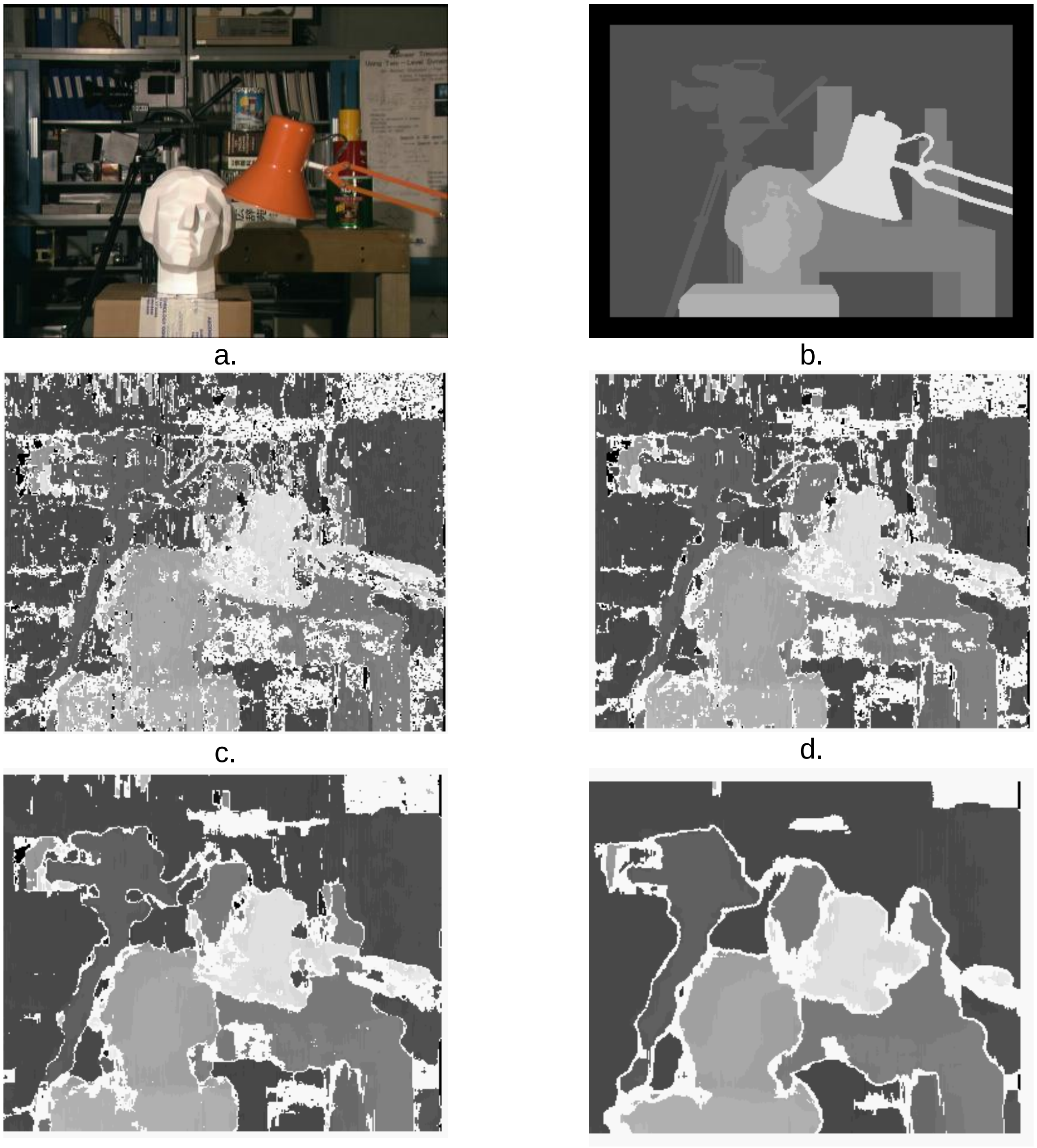
Stereo correspondence algorithms can be classified as local or global approaches. Local approaches use support windows to measure distances between the pixels on the input images. Global approaches are based on the minimization of an energy equation, where smoothness assumptions are modeled. (Scharstein & Szeliski, 2002) presents a pipeline commonly followed by stereo correspondence algorithms according to the presented taxonomy. The presented steps are listed as follows:

1. Matching cost
2. Cost (support) aggregation
3. Disparity computation / optimization
4. Disparity refinement

The matching cost step computes the cost of assigning different disparity hypotheses to different pixels. An evaluation of this matching cost functions can be found in (Heiko Hirschmuller & Scharstein, 2007). In the cost aggregation step the initial matching costs are aggregated spatially over support regions. Next, the best disparity hypothesis for each pixel is computed. Finally, the estimated disparity maps are processed to improve mismatches or fill pixels without disparity value assigned (Scharstein & Szeliski, 2002). The next section summarizes the main features of stereo correspondence algorithms.

- **Local Methods**

In a local stereo correspondence approach the disparity values are estimated independently from other disparities. The emphasis in local methods is on the matching cost and cost aggregation steps. This methods use the winner-take-all (WTA) optimization for disparity computation, where at each pixel the best (lower cost) disparity hypotheses is chosen. Conventional local methods rely on distance metrics on fixed windows for matching cost computation. Selecting the right window is important, since windows must be large enough to contain sufficient texture and yet small enough so that they do not straddle depth discontinuities (Szeliski, 2010). Figure 2.2.3.1 shows the results of disparity map estimation using a simple block matching algorithm with a fixed squared window from opencv (Bradski & Kaehler, 2008) employing SAD as metric for matching cost computation.



a. b.

c. d.

e. f.

Fig. 2.2.3.1. Left image for tsukuba (a), *ground truth* disparity map (b), disparity estimation computed with a local algorithm using SAD and a fixed window of sizes 5x5 (c), 7x7 (d), 15x15 (e), 21x21 (f) (Scharstein & Szeliski, 2002).

The mentioned drawbacks for fixed windows arise the motivation to develop local stereo correspondence approaches using adaptive support regions. The proposed approaches include: Methods based on multiple or shiftable support regions, where multiple symmetric square windows centered at different locations are used to aggregate the matching cost; Methods based on adaptive window size or shape, where a different support window is computed for each pixel; and methods based on adaptive weight, where the influence of each pixel during the disparity estimation process is computed (I. Cabezas, 2013; Scharstein & Szeliski, 2002).

- **Global Methods**

A global stereo correspondence method performs some optimization or iteration steps after the disparity computation phase. These methods are commonly formulated in an energy minimization framework and often skip the cost aggregation step, since smoothness assumptions are included in the energy minimization model (Szeliski, 2010). The goal is to minimize the global energy of:

$$E(d) = E_d(d) + \lambda E_s(d). \tag{7}$$

Here, E_d is the data term measuring how well the disparity function agrees with the input image pair as follows:

$$E_d(d) = \sum_{(x,y)} C(x, y, d(x, y)) \tag{8}$$

Where C is the matching cost function. And E_s is the smoothness term encoding smoothness assumptions, often restricted only to measure the differences between neighboring pixels' disparities (Szeliski, 2010).

$$E_s(d) = \sum_{(x,y)} \rho(d(x, y) - d(x + 1, y)) + \rho(d(x, y) - d(x, y + 1)) \tag{9}$$

Where ρ is an increasing function of disparity difference.

Dynamic programming, graph cuts and belief propagation are widely adopted for energy minimization in stereo correspondence global approaches. Dynamic programming minimizes the energy equation for independent scanlines in polynomial time, which commonly leads to streaking artifacts. The graph cuts approach states the energy minimization problem as the process of finding a minimum cut in a graph, while the belief propagation strategy solves the problem by iteratively sending messages between four connected neighboring nodes (pixels) on the image (I. Cabezas, 2013; Szeliski, 2010). In global stereo correspondence methods additional terms can be used for penalizing occlusions, among others.

In order to test the proposed method in these thesis two algorithm implementations available in the opencv library will be used. Firstly, the block matching algorithm (BM) is a local stereo correspondence algorithm that measures similarity between image regions (blocks) to estimate disparity. Initially, a reference block is defined in the reference (left) image, surrounding a point where the disparity will be estimated. Then, the sum of absolute differences (SAD) is computed for this block and compared against the SAD of the horizontal neighbors in the right image. Finally, the disparity is computed as the relative displacement between reference block in the left image and the block in the right image with the closest SAD (Bradski & Kaehler, 2008).

The semi-global block matching algorithm (SGBM) implementation in opencv is a modification of the stereo correspondence algorithm proposed in (H. Hirschmuller, 2008). Here, an energy equation optimization is performed in a similar way as a global stereo approach. The main difference is that the Energy minimization is performed along individual 1D paths instead of the regular 2D global minimization for a pixel *P* as shown in fig 4.1.1.1.

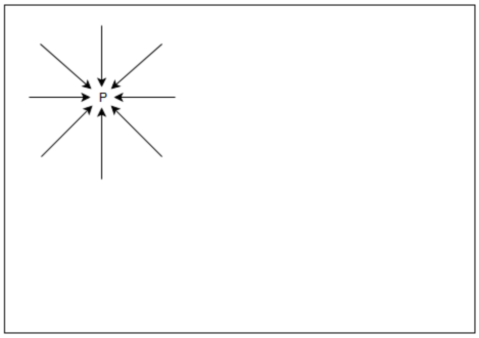


Fig. 4.1.1.1. SGBM approach

According to the opencv documentation and in the context of this thesis, the implemented algorithm differs from the original as follows: Firstly, considers five 1D paths instead of eight by default. Secondly, the algorithm match blocks instead of individual pixels, however, the block size parameter can be set to 1. Thirdly, the mutual information cost function is not implemented. Instead, a simpler Birchfield-Tomasi sub-pixel metric from (Birchfield & Tomasi, 1998) is used. A post-processing speckle noise reduction step is performed.

2.3 Image quality metrics

Image Quality assessment is an active area of research with an important role in several image processing applications. Several metrics have been proposed to develop an objective assessment well correlated with perceived human quality measurement or subjective methods. The objective Image Quality Assessment approaches can be classified into full-reference, reduced-reference and no-reference (Bhola, Sharma, & Bhatnagar, n.d.). For this work a comparison between synthesized right views and reference right view is required. Hence, a full reference image quality assessment is used. For this purpose two classes of image quality metrics are available: statistical error metrics and human visual system feature based metrics (HVS). The most widely statistical error metrics used in full-reference image quality are the MSE and PSNR. These metrics are simple to compute and have a mathematical clear meaning but not well matched to perceived visual quality (Z. Wang, Bovik, Sheikh, & Simoncelli, 2004). These metrics are defined as follows:

Statistics error metrics:

MSE: Standing for mean squared difference is the Euclidian distance between the original and the degraded images, is defined as:

$$MSE = \frac{1}{M \times N} \sum_{i=1}^{M} \sum_{j=1}^{N} (x_{ij} - y_{ij})^2 \qquad (10)$$

Where x_{ij} is the value of the image pixel located in the coordinates (i, j) and M, N are the dimensions of the compared images.

PSNR: The Peak signal to noise ratio is a well-known index defined as the ratio between the maximum possible power of a signal and the power of corrupting noise that affects the fidelity of its representation (Bhola et al., n.d.). PSNR can be defined as:

$$PSNR = \frac{10log_{10} \times \ 255^2}{MSE} \qquad (11)$$

Where 255 is the maximal possible value the image pixels when pixels are represented using 8 bits per sample.

AD: The average difference is simply the average of difference between the reference and the test images, given by the equation (12):

$$AD = \frac{1}{M \times N} \sum_{i=1}^{M} \sum_{j=1}^{N} (x_{ij} - y_{ij}) \qquad (12)$$

MAE: The mean absolute error is the average of the absolute difference between the reference and test images:

$$MAE = \frac{1}{M \times N} \sum_{i=1}^{M} \sum_{j=1}^{N} |x_{ij} - y_{ij}| \qquad (13)$$

MD: The maximum difference is the maximum absolute difference between the reference signal and test image, defined as:

$$MD = \max |x_{ij} - y_{ij}| \qquad (14)$$

PMSE: The Peak Mean Square Error It is given by the following equation:

$$PMSE = \frac{1}{M \times N} \times \frac{\sum_{i=1}^{M} \sum_{j=1}^{N} (x_{ij} - y_{ij})^2}{(\max(x_{ij}))^2} \tag{15}$$

Human visual system (HVS) feature based metrics:

SSIM: The structural similarity index is a metric designed to improve on traditional methods like PSNR and MSE in image quality assessment. This metric compares two images using information about luminous, contrast and structure using local windows. The measure between two local windows x and y of common size is given as (Z. Wang et al., 2004):

$$SSIM(x,y) = \frac{\{(2\mu_x\mu_y + C_1)(2\sigma_{xy} + C_2)\}}{\{(\mu_x{}^2 + \mu_y{}^2 + C_1)(\sigma_x{}^2 + \sigma_y{}^2 + C_2)\}} \tag{16}$$

Where μ_x is the average of x; μ_y is the average of y; σ_x, σ_y are the standard deviations between the original and processed images pixels respectively. $C1, C2$ are positive constants chosen empirically to avoid the instability of measure.

MSSIM: The mean of SSIM is known as mean structural similarity index metric and it is given as:

$$MSSIM(X,Y) = \frac{1}{M} \sum_{i=1}^{M} SSIM(x_i, y_i) \tag{17}$$

Where X and Y are the assessed and reference images and M is the total of the (x_i, y_i) local windows assessed. For images of very different quality which have roughly same mean square error, with respect to the original image. MSSIM gives a much better indication of image quality (Z. Wang et al., 2004).

3. Literature review

This section presents the review findings in the fields of interest using a systematic review approach. The following subsections describe the systematic review approach and its results for stereo correspondence algorithms and image synthesizing.

3.1 Systematic review

A systematic review an approach to identify, evaluate and interpret all available information relevant to a particular research question, topic or phenomenon of interest. This research method must be performed in accordance to a predefined strategy which must allow measuring the completeness and quality of the review (Kitchenham, 2004). The systematic review method is inspired from medical research and has started to get attention lately in software engineering. Briefly, a systematic review goes through existing researches reviewing them in-depth and describing their methodology and results (Petersen, Feldt, Mujtaba, & Mattsson, 2008).

Systematic reviews have several advantages and disadvantages compared to regular literature reviews. Some benefits of systematic reviews include the bias reduction through a well-defined research strategy, a wider detection range of studies and thus, more general conclusions. Systematic reviews also has several drawbacks, with the considerable effort it requires compared to regular literature reviews being the main one (Petersen et al., 2008).

Protocol. The protocol specifies the steps that are going to be followed in order to perform the systematic review. A pre-defined protocol is necessary to reduce the possibility researcher bias. In medicine, review protocols are usually submitted for a peer review (Kitchenham, 2004). The following elements are used in this work to define the systematic review protocol:

Research questions: Research questions definition is the first step when conducting the systematic. The research questions are related to the concerns that should be answered during the review.

Study selection: The study selection includes three elements: keywords/search string definition, sources selection and inclusion/exclusion criteria. Keywords allow determining the search string that is going to be used on the web search engines in order to find studies of interest for the review. The sources selection defines the databases, journals, and conference proceedings are going to be searched. Finally, the inclusion/exclusion criteria are intended to identify those studies that provide direct evidence about the research question.

Results summary: In this work a brief summary of results for the performed systematic review is presented. The purpose of this section is to explicitly show possible trends in the researched field.

Findings on the area: In this section a synthesis of the selected studies is performed. The synthesis must collect all the information needed to address the review questions. Forms are commonly used to fulfil this component. In this work a descriptive synthesis is presented for the systematic reviews performed.

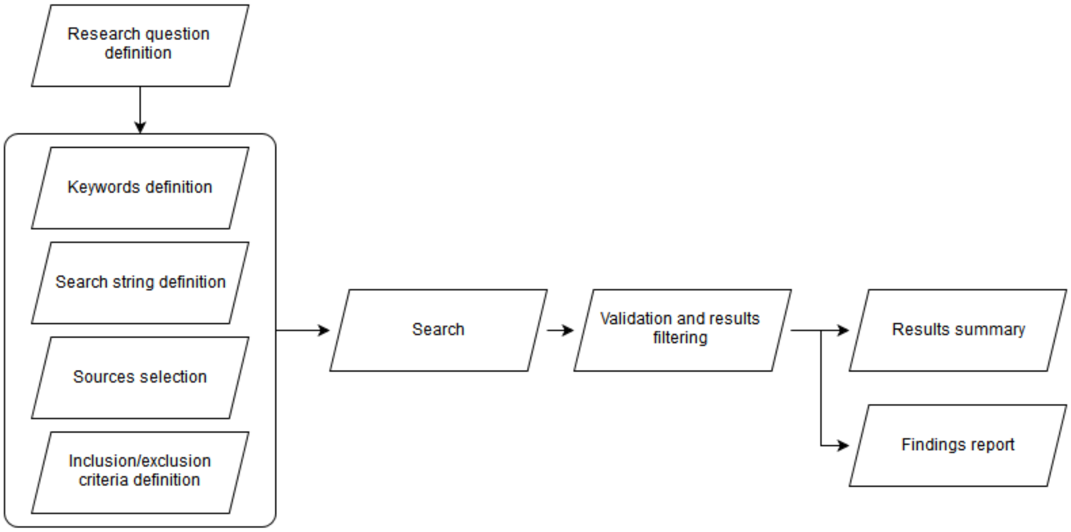

Research question definition
Keywords definition
Search string definition
Sources selection
Inclusion/exclusion criteria definition
Search
Validation and results filtering
Results summary
Findings report

Fig 3.1. shows a flow diagram for conduction the systematic review in the context of this thesis.

3.2 Stereo Correspondence Evaluation Methods

3.2.1 Systematic review

Research questions. This systematic review is based on two research questions:

(i) Which are the evaluation methods and evaluation frameworks for assessing the quality of disparity maps obtained from stereo correspondence algorithms?

(ii) Which are the stereo image datasets available to perform quality assessment of stereo correspondence algorithms?.

Study selection. To define the study selection approach three elements are used in this work: keywords/search string definition, sources selection and inclusion/exclusion criteria.

Keywords and search string. Defined keywords were classified in stereo correspondence related terms and quality related terms as follows:

Stereo correspondence related terms: *stereo matching, stereo correspondence, stereo algorithm, stereo vision and disparity map*.

Quality related terms: *evaluation, measure, quality, metric, assessment and performance.*

Based on keywords previously defined , the search string below was used:

(("stereo" AND "matching") OR ("stereo" AND "correspondence") OR ("stereo" AND "algorithm") OR ("stereo" AND "vision") OR ("disparity" AND "map") OR ("Stereoscopic" AND "image")) AND (("evaluation") OR ("measure") OR ("quality") OR ("metric") OR ("assessment") OR ("performance"))

Sources selection*.* Information sources are selected according to the defined research question. Among multiple information sources, bibliographic databases have high reliability. The Scopus database was chosen since it integrates important digital libraries addressing visual computing topics. A total of 5937 papers were obtained as searching results.

inclusion and exclusion criteria. Inclusion criteria allows to consider specific studies, whilst exclusion criteria filters out obtained results not closely related to research questions. The inclusion/exclusion criteria for this review were defined as follows:

(i) the study should approach a method, strategy, metric or dataset for assessing quality of disparity maps or stereo correspondence algorithms.

(ii) Stereo image evaluation for comfort measuring or 3DTV applications that does not include assessment of disparity maps will be excluded.

(iii) The study publication date should be equals or greater than 2005

Additionally, the use of control papers allows to quickly verifying the coherence between search string and obtained results. This requires of some background on the addressed topic. (I. Cabezas et al., 2012; H. Hirschmuller & Scharstein, 2009; Morales & Klette, 2009, 2011) were defined as control papers.

Results summary. Categorization of results allows constructing a visual summary, indicating trends. Figure 3.2.1.1 shows the quantity of published papers per year. Figure 3.2.1.2 shows the quantity of published papers by trend, classified as papers proposing or applying strategies without ground-truth (confidence metrics and prediction error approaches), stereo datasets and ground-truth methods.

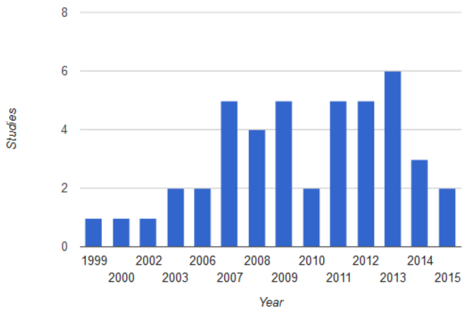


Fig. 3.2.1.1. Quantity of published studies per year.

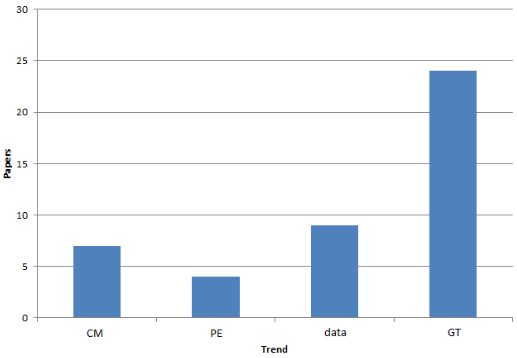


Fig. 3.2.1.2. Quantity of published studies by trend, Confidence measures, prediction error, datasets, ground-truth methods, respectively.

3.2.2 Findings on the area

Evaluation methods can be classified into two main approaches: evaluation approaches using ground-truth data and evaluation approaches in the absence of disparity ground-truth data, summarized in the next two subsections.

- **Evaluation methods using ground truth**

Evaluation methods were firstly proposed in order to measure the improvement on the stereo correspondence field. The Middlebury dataset and method is presented in (Scharstein & Szeliski, 2002). This method allows intra-technique and inter-technique evaluation of stereo correspondence algorithms. The dataset introduced on this method is available at Middlebury's website, including several stereo images and ground-truth data, still active on its third version. This method measures the estimated disparity map quality using the BMP and RMS metrics against ground-truth data. Different error criteria are associated to image segments: all, the entire image; nonocc, areas that are not-occluded;disc, areas near depth discontinuities and occluded regions; and textureless, areas of low texture.

In (Szeliski & Zabih, 2000) an evaluation is performed using two separate approaches: a comparison against ground-truth data and the prediction error approach. This work defines an error as an estimation disagreeing from the ground-truth disparity value in more than 1 pixel. An error criterion is used in order to measure the algorithm's performance under different situations such as occluded pixels or low texture regions.

In (Haeusler & Klette, 2010) is pointed out that it might be possible to quantify the quality of recorded stereo images with respect to some measures, which may be used for indicating domain of relevant scenarios when performing evaluations for some particular test data. The aim of the work is to judge the complexity of a specific stereo dataset and its qualitative relation to other datasets.

Robustness to radiometric changes and noise between views is required in stereo correspondence algorithms for real world applications. In (H. Hirschmuller & Scharstein, 2009) all possible combinations of 13 cost function and three stereo correspondence algorithms are evaluated. Cost functions include absolute difference, the sampling-insensitive absolute difference, and normalized cross correlation, as well as their zero-mean versions. The stereo correspondence algorithms are local, semiglobal and global approaches. The study measures the performance of all costs combinations in the presence of simulated and real radiometric differences, including exposure differences, vignetting, varying lighting, and noise. The Middlebury dataset is used on this work and performance measures are done using BMP metric. In the same way in (Leclercq & Morris, 2003) the robustness to noise is measured for stereo correspondence using the Middlebury dataset and the SMR metric.

In (Sellent & Wingbermühle, 2012) a quality assessment of stereo correspondences based on histogram differences is proposed. The improvement of this study is based on the idea of assessing when an object is missed from a non-dense disparity map. The proposed method divides the image in small subregions where disparity histograms are calculated. For each region the histogram distances are calculated using the earth mover’s distance (EMD) and averaged.

(Kondermann et al., 2015) proposes a method to create arbitrary stereo ground truth datasets with reliable per-pixel error bars and a method to add error bars to image sequences with disparity ground truth. It is based on previously measured point clouds and arbitrary calibrated cameras and therefore versatile for indoor as well as outdoor applications.

An evaluation method for parameter setting is proposed in (Kostlivá, Čech, & others, 2007). It considers two error types: the error rate and the sparsity rate, for accuracy and completeness measuring respectively. These error definitions are based on four principles: orthogonality, symmetry, completeness and algorithm independence.

A cluster ranking intra-technique evaluation method is proposed in (Neilson & Yang, 2008). The proposed method consists on using a statistical inference technique (ANOVA) to rank the accuracy of disparity estimation algorithms combining ranks from rom multiple stereo pairs. Imagery used in this study includes 90 synthetic images, with three different levels of noise, generated by a ray tracing method and 18 images from the Middlebury's image repository, some of them having radiometric changes. The BMP measure is used, only, according to the nonocc error criterion.

The R-SSIM measure is proposed in (Malpica & Bovik, 2008). The R-SSIM is a modification of the Multi-scale Structural Similarity index (MS- SSIM). The obtained results by using R-SSIM measure are statistically correlated to obtained results from BMP measure. Nevertheless, the final ranking assigned to disparity estimation algorithms, using the evaluation model of the Middlebury method, varies considerably when the R-SSIM measure is used.

In (Yinghua Shen, Chaohui Lu, Pin Xu, & Lili Xu, 2011) the SSIM and PSNR measures are compared for disparity maps with added salt and pepper noise. The authors conclude that obtained PSNR values are closer to the scores assigned by subjective evaluation.

(I. Cabezas et al., 2012) proposes a quality metric for disparity map using ground truth data. Bad matched pixels (BMP) is a widely used metric for disparity ground truth comparison but this measure ignores the inverse relation between depth and disparity. Also, using BMP small errors are counted the same way than a large errors. therefore, two disparity maps with equal BMP percentages may produce different 3D reconstructions. The proposed BMPRE metric offers a clear and concise interpretation of a disparity estimation error considering both the error magnitude and the inverse relation between depth and disparity.

An evaluation involving estimation accuracy and computational efficiency is proposed in (vanderMark & Gavrila, 2006). The imagery test-bed used includes the Middlebury's data set and a dataset termed Lab, acquired in uncontrolled environments using an off-the-shelf stereo camera. This proposal is focused on disparity estimation algorithms suitable to be used in application domains requiring near real-time performance and/or to be executed on hardware platforms with limited resources. The complement of the BMP measure is used to gather errors according to the nonocc and the disc criteria.

Several evaluation methods oriented to specific contexts are proposed or applied in the stereo correspondence field; these studies are summarized in table 3.2.2.1.

Application / Context	**Reference**
Autonomous vehicle applications	(Geiger, Lenz, & Urtasun, 2012; Hamilton, Breckon, Bai, & Kamata, 2013; Leibe, Cornelis, Cornelis, & Van Gool, 2007; Morales & Klette, 2011; Morales, Vaudrey, & Klette, 2009; Steingrube, Gehrig, & Franke, 2009; vanderMark & Gavrila, 2006)
Face reconstruction	(Woodward, Leclercq, Delmas, & Gimel'farb, 2006)
Real time oriented evaluation	(Gong, Yang, Wang, & Gong, 2007; Tombari, Mattoccia, & Di Stefano, 2010)
Agriculture applications	(Nielsen et al., 2007)
Pedestrian detection	(Keller et al., 2011; Philip Kelly, 2007; P. Kelly, O'Connor, & Smeaton, 2008)
Silicon retina stereo cameras	(Kogler, Eibensteiner, Humenberger, Gelautz, & Scharinger, 2013)
Remote sensors	(Aguilar, del Mar Saldana, & Aguilar, 2014)

Table 3.2.2.1. Stereo correspondence evaluation methods oriented or applied to specific contexts.

- **Evaluation methods without ground truth**

Evaluation methods that do not use ground truth data can be classified as prediction error approaches and confidence measure approaches (Morales & Klette, 2011; B.-S. Shin et al., 2015). The prediction error approach (Szeliski, 1999) suggest the prediction of a novel view of the scene. The predicted view can be compared to a reference view obtained from a third camera in a known position. However, error scores reflect not only the accuracy of the disparity estimation algorithm, but also the accuracy of the selected rendering algorithm, since the rendering process of the predicted view has to deal with interpolation or extrapolation issues (Scharstein & Szeliski, 2002; Sellent & Wingbermühle, 2012). Confidence metrics are used to measure the reliability of the estimated disparity value for each pixel (Morales & Klette, 2011). Several stereo correspondence algorithms use confidence metrics as part of their estimation processes in order to refine the resulting disparity maps.

In (Morales & Klette, 2009) three stereo correspondence algorithms are evaluated using the prediction error approach for an autonomous navigation context. The evaluation is performed using synthetic data from (Wedel et al., 2008). In this work, a reference image is acquired using a third camera and a projected view is generated from estimated disparity maps for each algorithm. The reference and predicted images are compared using root mean squared (RMS) and normalized cross correlation (NCC). The ranking obtained for the evaluation method resembles a ranking obtained using ground truth data comparison.

More sophisticated metrics can be applied to compare the predicted and reference images in the prediction error approach. In (Fuhr et al., 2013) the authors use a prediction error approach applied to the view interpolation problem. Again, a reference third view and predicted view are compared. This study uses structural similarity index (SSIM) and peak signal to noise ratio (PSNR) measures to calculate a quality metric. Study results shows low correlation between the traditional ground truth based evaluation method using BMP and the proposed view interpolation metrics.

Recently, in (Vandewalle & Varekamp, 2014) an evaluation method for stereo video sequences is proposed. The authors present a two-dimensional analysis of disparity map quality using a matching error and a temporal instability metrics. The matching error measure is a prediction error based approach, but particularly in this work the evaluation does not require a third view. Instead, the evaluation method predicts the right view using the left view and the estimated disparity map. The reference and predicted right images are compared using the mean absolute difference (MAD) metric. Finally, the temporal error is calculated using motion estimations, where disparity maps with high temporal instability will lead to a higher temporal error.

Confidence metrics are commonly used as a supporting step on stereo correspondence algorithms and can also be used as a quality metric. A quantitative and qualitative comparison for confidence metrics is presented on (Xiaoyan Hu & Mordohai, 2012). Confidence metrics are expected to be high for correct disparities and low for errors, detect occluded pixels and useful to select between several disparity hypotheses. In this work, the evaluation is performed by comparing the confidence measures against the disparity maps errors producing ROC curves. The confidence measures are classified according to the aspects of cost they consider; those aspects include local properties of the cost curve, local minima of the cost curve, consistency between left and right disparity maps among others. Finally the study shows a detailed performance analysis is presented where advantages and disadvantages for each metric are mentioned.

A classifier using confidence measures as input features for stereo matching refinement is proposed in (Varekamp, Hinnen, & Simons, 2013). The proposed stereo correspondence algorithm is supported by an AdaBoost approach, where estimated disparities are classified as either ‘correct’ or ‘incorrect’. The feature vector used on the classifier includes confidence metrics such as average color components, texture, color variation, disparity variation, among others.

In the same way, (Haeusler, Nair, & Kondermann, 2013) proposes the use of confidence metric as features for a random decision forest classifier. This study is developed using stereo images and ground truth data from the KITTI dataset (Geiger et al., 2012). The confidence metrics used as features include Entropy of disparity costs, peak ratio measure, consistency between left and right disparity, horizontal gradient, among others. Learning samples are categorized into two classes: good and bad disparities. This work shows that a classifier using confidence measures can be an appropriate approach to increase accuracy in stereo error detection.

A quality metric for depth maps using unsupervised no reference segmentation quality metrics is proposed in (Milani, Ferrario, & Tubaro, 2013). The quality metric is calculated by checking consistency between a segmented depth map and one input image of the same scene. The results show some correlation degree between the proposed quality metric and the prediction error approach where a MSE metric is used to compare the predicted and reference views.

A correlation assessment between the prediction error method and 2D image metrics is performed in (B.-S. Shin et al., 2015). The assessment is done for stereo video sequences under absence of ground truth (Hermann, Morales, & Klette, 2011). The tree proposed data measures are called SL, SS and LR, dealing with image homogeneity, standard deviation of the Sobel image and similarity between stereoscopic images, respectively. Authors propose the SL or LR measures to be tested in order to replace the dependence on a third view or to combine them with the third eye method to achieve a robust evaluation approach.

3.2.3 Discussion

According to the findings on this systematic review a taxonomy of the different state-of-art methods for assessing stereo correspondence algorithms is introduced in this work as shown in Figure 3.2.3.1.

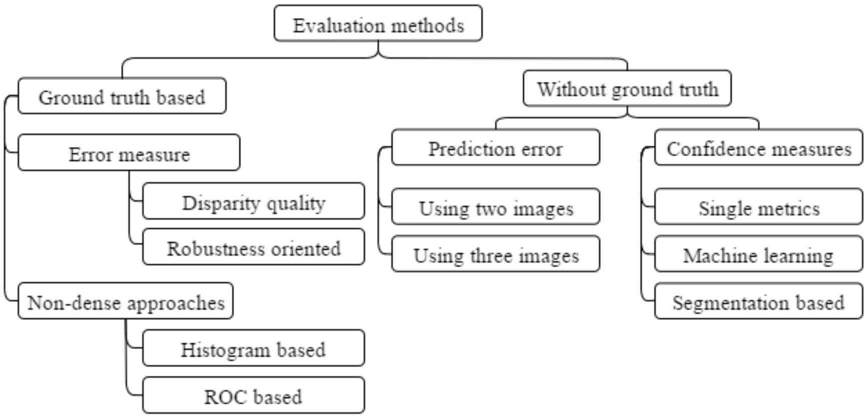


Fig. 3.2.3.1. Disparity map evaluation methods taxonomy.

Middlebury method is one of the most used ground-truth based evaluation methods. In Middlebury's method BMP and RMS are used as metrics (Scharstein et al., 2014; Scharstein & Szeliski, 2002, 2003). Several metrics including SSIM, PSNR (Yinghua Shen et al., 2011), R-SSIM (Malpica & Bovik, 2008), BMPRE (I. Cabezas et al., 2012), SZE (I. Cabezas et al., 2011), disparity gradient and disparity acceleration (Zhang, Hou, Shen, & Yang, 2009) have been also proposed in order to estimate the disparity quality. Nevertheless, there is a lack of consistency on the evaluation results achieved by considering different error measures (I. Cabezas, 2013).

Robustness to noise and radiometric changes are addressed in (H. Hirschmuller & Scharstein, 2009) and (Leclercq & Morris, 2003). These approaches consider radiometric changes artificially generated on the Middlebury datasets and artificial noise on a synthetic dataset, respectively.

Ground-truth based proposals also include histogram (Sellent & Wingbermühle, 2012) and ROC (Kostlivá et al., 2007) based evaluations, where the sparsity of estimated disparity maps its handled explicitly. The histograms approach is focused on disparity distribution and outliers. The ROC approach is focused on studying a wide range of parameter settings for a single algorithm based on the defined error and sparsity rates.

Regarding to evaluation methods in the absence of ground-truth data, prediction error is proposed in (Szeliski, 1999). This approach requires the use of a third camera and therefore the modification of the standard stereo acquisition system. In (Vandewalle &

Varekamp, 2014) the prediction error method is performed using the two standard stereo images, removing the additional work at the acquisition stage.

According to (Xiaoyan Hu & Mordohai, 2012), confidence metrics are grouped as matching cost metrics, local properties of the cost curve, entire cost curve metrics, consistency between the left and right disparity maps and distinctiveness based confidence measures. Confidence measures can be used as input features for classifiers as is presented in (Varekamp et al., 2013) and (Haeusler et al., 2013). The confidence measure proposed in (Milani et al., 2013) is calculated by checking consistency between a disparity based segmentation against a color based segmentation of a view of the stereo image pair used as input. This approach is limited by the assumption of smooth disparity changes over color based segments.

Datasets for stereo correspondence algorithms evaluation include the Middlebury (Scharstein & Szeliski, 2002), KITTI (Geiger et al., 2012) and the Enpeda Image Sequence Analysis Test Site (EISATS) (Wedel et al., 2008), where several stereo images with their respective ground-truth are available. Additionally, methods to create and compare datasets are discussed in (Kondermann et al., 2015) and (Haeusler & Klette, 2010) respectively.

Although the progress on the stereo correspondence problem can be qualitatively inferred, for instance, by the application of different optimization strategies, or by the approaches proposed on different aspects of the disparity estimation process, an objective and quantitative assessment is required not only to determine if a particular algorithm can be considered as superior to other or others -within a particular context- , but also, in order to properly provide feedback to the researcher or practitioner. In this sort of ideas, this paper may result interesting to the reader for two main reasons: by the particular findings on the stated question, and highlighting how a systematic review can be used on visual computing research (Vargas, Cabezas, & Branch, 2015). Table 3.2.3.1 shows a technique comparison of the state-of-art prediction error approaches found on the systematic review and the proposed method on this thesis.

Proposal	**# images required**	**Interpolations / extrapolations**	**Disparity map preprocessing**	**Metrics**	**Video sequences**
(Szeliski, 1999)	3	Yes	No	-	No
(Morales & Klette, 2009)	3	Yes	No	RMS	No
(Fuhr et al., 2013)	3	Yes	No	SSIM – PSNR	No
(Vandewalle & Varekamp, 2014)	2	Yes	Yes	MAD	Yes
Proposed	2	No	No	MSSIM - MSE	No

Table. 3.2.3.1. Prediction error approaches.

3.3 Image synthesis

3.3.1 Systematic review

Research question. This systematic review is based the following research question:

(i) Which are the state-of-art algorithms or approaches that can be used to synthesize views of a scene acquired with stereo vision.

Study selection. To define the study selection approach three elements are used in this work: keywords/search string definition, sources selection and inclusion/exclusion criteria.

Keywords and search string. Defined keywords were classified in stereo correspondence related terms and image synthesis related terms as follows:

Stereo correspondence related terms: *stereo matching, stereo correspondence, stereo vision, disparity map and stereoscopic image*.

Image synthesis related terms: *stereo reprojection, view synthesis, warping, reconstruction, DIBR.*

Based on keywords previously defined, the search string below was used:
(("stereo matching") OR ("stereo correspondence") OR ("stereo vision") OR ("disparity map") OR ("stereoscopic image") OR ("stereo reprojection")) AND (("view synthesis") OR ("warping") OR ("reconstruction") OR ("DIBR"))

Sources selection. Information sources are selected according to the defined research question. Among multiple information sources, bibliographic databases have high reliability. The Scopus database was chosen since it integrates important digital libraries addressing visual computing topics. A total of 1036 papers were obtained as searching results.

inclusion and exclusion criteria. The inclusion/exclusion criteria for this review were defined as follows:

(i) The study should approach a method, strategy, or algorithm for synthesizing views from a stereo image and its disparity map..

(ii) The study publication date should be equals or greater than 2010

Additionally, the use of control papers allows to quickly verifying the coherence between search string and obtained results. This requires of some background on the addressed topic. (Vandewalle & Varekamp, 2014) and (Fehn, 2004) were defined as control papers.

Results summary. Figure 3.3.1.1 shows the quantity of published papers per year. Figure 3.3.1.2 shows the quantity of published papers by trend, classified as studies performing image synthesis using image domain warping, layered approaches or DIBR.

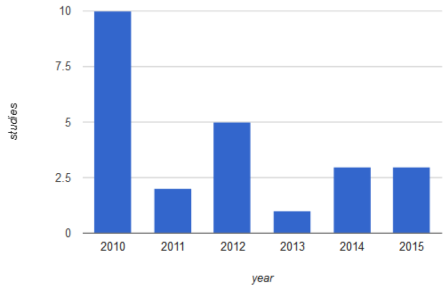


Fig. 3.3.1.1. Quantity of published studies per year.

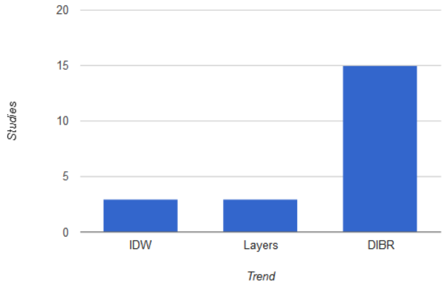


Fig. 3.3.1.2. Quantity of published studies by trend, Confidence measures, prediction error, datasets, ground-truth methods, respectively.

3.3.2 Findings on the area

Image warping approaches for view synthesis can be classified according to the systematic review into: image domain warping, layer based and DIBR approaches. Each one of the studies found is categorized and summarized in the following subsections.

- **Image domain warping**

In (Yao, Wang, Lin, & Zhang, 2015) three main contributions are proposed for stereo to multiview content generation. First, proposes an adaptive meshing that uses saliency into the image warping approach, which aims to reduce the computational complexity at the expense of slight decrease in quality. Second, a simple and effective method based on block matching algorithm to generate the sparse disparity map. And third, they manage to accelerate the algorithm’s execution speed with parallelization strategies on graphic processing units (GPUs). The algorithms proposed in this work include adaptive meshing to segment the image into blocks, sparse stereo correspondence based on block matching, energy equation construction based on the sparse disparity map and a virtual view rendering based on the energy equation. Here, the energy equation is oriented to compute the block warping; afterwards the remaining points within blocks are mapped from the input to the virtual views according to bilinear interpolation.

A warping-based method for synthesizing multiple views from a binocular stereoscopic image is presented in (Huang, Huang, Huang, Chen, & Chuang, 2012). This work proposes a non-dense disparity estimation based on feature matches to guide the image warping and synthesize novel views. The locations of the matched feature pairs are interpolated or extrapolated to estimate their corresponding coordinates in the desired virtual view. An energy function is proposed for image warping that includes the following three terms: matched feature correspondence, Content coherence and a Line preserving term. The energy function is minimized using standard sparse linear solvers.

(Kim, 2010) proposes an intermediate view synthesis method suitable for a rectangular multi-view camera system. This method uses three reference images from a 4x4 rectangular multi-view camera arrangement. First, the virtual view is divided in regions which are synthesized from the nearest image in the arrangement. Second, an edge-based feature extraction process is performed in order to compose a triangular mesh using Delaunay triangulation. Third, a mesh-based disparity estimation is performed, where a disparity is assigned for each triangle in the mesh. Fourth, the virtual view is synthesized by an affine transform using the reference views and estimated disparity maps. Finally, a hole filling and post-processing steps are performed on the synthesized image to reduce the disocclusions and visual artifacts. This method is evaluated using the PSNR metric.

- **Layer based approaches**

In (N.A. Manap & Soraghan, 2014) and (Nurulfajar Abd Manap & Soraghan, 2011) the authors propose an intermediate view synthesis method based on disparity estimation depth map layers. This approach is performed on two stages: stereo matching and view synthesis modules. In the first stage, disparity estimation through area based stereo matching algorithm is adopted to obtain the disparity depth map. A left-right consistency (LRC) check is performed to eliminate the half-occluded pixels in the final disparity map. In the second stage, the disparity map is divided into layers using the disparity histogram distribution. Each layer is warped according to the layer disparity. The final novel view synthesis obtained by blending and flattening the layers into a single image.

A multi-view stereoscopic image synthesis algorithm for 3DTV system using depth information and a texture image acquired using a depth camera is proposed in (Choi, Seo, Yoo, & Kim, 2013). The algorithm uses a parallel camera model and divides the images in foreground and background layers. Left and right stereo images are synthesized DIBR according to the disparity estimated for layers. A 4-neighbor pixels spatial interpolation algorithm that takes into account the direction of background objects' edges is proposed for hole filling.

- **DIBR**

Depth-image-based rendering (DIBR) is the process of synthesizing virtual views of a scene from still or moving color images and associated per-pixel depth information (Fehn, 2004). In (Liu, Zhang, Cui, & Ding, 2015) the shift-sensor approach is used in order to synthesize two images for a stereo system using a single centered image and its respective disparity map. Major contributions of this work are focused on disparity map pre-processing and hole filling. An enhanced adaptive directional filter is introduced for disparity map pre-processing; this filter can not only smooth sharp depth change, but also overcome the disocclusion problem while providing good, reasonable disparity cues. The proposed hole filling method is achieved by simply interpolating image of pixel information in the foreground and background for the input image which can lead to obvious visible disocclusion artifacts particularly on object boundaries with large size holes.

A disparity refinement method near to object boundaries for quality enhancement of the synthesized image is presented in (Lee & Yoo, 2015). A noisy disparity map is obtained using stereo matching. In order to improve the disparity map quality for view synthesis a consistency check between left and right disparity maps, occlusion detection processes and a disparity map refinement using a joint bilateral filter are performed. The proposed method is compared against 3 disparity refinement methods for view synthesis using PSNR.

In (Lei, Chen, & Shi, 2014) a new hole-filling algorithm based on pixel labeling is proposed. Left and right views are synthesized from a centered image and its respective disparity map using the shift-sensor approach. Hole pixels in the synthesized left and right images are filled according to the non-hole pixels in a eight-neighborhood. Holes, corners and sides are processed differently.

(Zhu, Li, & Yu, 2014) proposes a virtual view synthesis using a SAD matching cost for stereo correspondence. Once the disparity maps are estimated, the synthesis process is performed using the also called disparity compensation method which can be derived from the shift-sensor approach. The disparity compensation method simply shifts horizontally a pixel in an image according to its disparity. A pair of synthesized images is obtained from left and right input images and the disparity map. Then, a hole filling process is then performed and the resulting images are then blended into the final synthesized view. PSNR and UIQI metrics are used to objectively evaluate the proposed method performance.

In (Riechert, Zilly, Müller, & Kauff, 2012) a rendering algorithm for DIBR, which uses a two-step rendering is proposed. First, a forwards mapping is applied to the disparity map. As soon as the disparity map is rendered to its new camera position, a backwards mapping of the virtual image's pixels becomes possible and the virtual image can be rendered in a second step. This enables the method to use sophisticated interpolation filters for the color values of each target pixel. Standard nearest neighbor, linear interpolation and a Lanczos3 filters are implemented in this work to synthesize the virtual view.

(Hsiao, Cheng, Wang, & Yeh, 2012) proposes a new algorithm that performs parallel warping and hole-filling operations, so the overall computation latency is significantly reduced. To achieve the parallelism between warping and hole filling a hole check is performed every time a pixel is shifted. Also, this method implements an overwrite logic in case of occlusions when warping, where the nearest pixel to the cameras is selected. Additionally a method called "raised disparity around edge" is performed in order to eliminate visual artifacts around edges.

An algorithm to generate content for multiview autostereoscopic displays is presented in (Geetha Ramachandran & Markus Rupp, 2012). First, two candidate intermediate views are generated from each one of the stereo views. The stereo views used are required to be rectified images. Hence, it can be assumed that the correspondences between points in the images occur along horizontal lines. The position of the pixels in the candidate new views is determined by shifting the pixels by scaled disparities. Then, the two candidate intermediate images and disparity maps are merged by placing pixels from both images into the new view and retaining pixels with greater disparity where pixels from both images occur at the same position. Finally, the proposed method is evaluated using the PSNR and SSIM metrics.

In (I.-Y. Shin & Ho, 2012) a method for real-time disparity estimation and intermediate view synthesis from stereoscopic images is proposed. In order to synthesize virtual viewpoint images a disparity map at the virtual viewpoint is estimated using hierarchical belief propagation. Then the synthesized view is estimated using a backward warping process. Holes on the disparity map and synthesized view are filled with neighboring pixels using an occlusion map. The resulting images are evaluating using the PSNR metric.

In (L. C. Tran, Pal, & Nguyen, 2010) and (L. Tran, Khoshabeh, Jain, Pal, & Nguyen, 2011) a method to synthesize intermediate views from two stereo images and their respective disparity maps is presented. The proposed method builds two placement matrixes for left and right images using the disparity maps. A placement matrix is a sparse matrix that contains 0 or 1 for each element to indicate that a pixel in the reference view is placed in in a specific coordinate in the virtual view. Using the placement matrixes each pixel in the virtual view is labeled as stable, unstable or occluded, where stable pixels have only one candidate pixel, unstable pixels have multiple pixel candidates and disocclusion pixels have no candidate pixel. The candidate pixels are obtained by shifting coordinates according to the disparity map. Occluded pixels are obtained using mean-shift segmentation and thresholding disparity on segments, where pixels with disparity that exceed ±20 from the mode are labeled as occluded. In the case of unstable pixels the candidate closer to cameras is picked. In (L. C. Tran et al., 2010) a discriminative CRF model is used to fill disocclusions. In (L. Tran et al., 2011) disoccluded pixels are filled with a exemplar-based image in-painting technique. An objective evaluation is conducted using the Middlebury dataset and the PSNR and SSIM metrics.

A view synthesis method which detects and then smooth out artifacts by anisotropic diffusion is proposed in (Devernay & Peon, 2010). First two disparity maps for the virtual view are estimated from left and right images. The intensity values on the synthesized image are interpolated from the warped left and right images using the virtual view disparity maps. Finally an artifact detection using a confidence map is proposed. This confidence map is used in the artifacts removal process with an anisotropic diffusion blurring.

A reliability model from epipolar geometry where the view interpolation algorithm is generated with the criterion of Least Sum of Squared Errors (LSSE) is proposed in (Yang, Yendo, Tehrani, Fujii, & Tanimoto, 2010a). The proposed algorithm can be considered as a reliable version of the conventional linear view blending. The proposed method is evaluated using PSNR metric. In (Yang, Yendo, Tehrani, Fujii, & Tanimoto, 2010b) and

(Yang, Yendo, Panahpour, Fujii, & Tanimoto, 2010) the reliability model is improved as a probabilistic model. These works propose a method for the plausible view synthesis of Free-viewpoint TV (FTV), using two input images and their depth maps. The main contributions of this work are the probabilistic model inferred for view synthesis and the reliability-based framework which adaptively synthesizes the virtual view. The probabilistic reliability model is proposed to guide the view interpolation. The view synthesis framework is dependent of the disparity error estimation which is performed using left-right disparity crosscheck. More accurate error approximation in the reliability computation would lead to better synthesis results. This method is evaluated using the PSNR metric.

(Devernay & Duchêne, 2010) applies image synthesis using baseline and viewpoint modifications. The proposed method is called hybrid disparity remapping and is a mixed technique between baseline and viewpoint modifications, preserving the global visibility of objects in the original viewpoints, but does not produce depth distortion or divergence. Baseline modification is a technique that generates a pair of new views as if they were taken by cameras placed at a specified position between the original camera positions; this technique is performed using the shift-sensor approach. Viewpoint modification is a similar technique that also allows changing the distance to screen but generally produces greater disocclusions. The transformation is evaluated on the tsukuba stereo image from Middlebury dataset.

A virtual view synthesis method is proposed in (Lü, Wang, Ren, & Shen, 2010), based on disparity map and image interpolation. Firstly, an initial disparity map of input stereo image is estimated using a stereo matching method based on adaptive weight. Next, occluded regions are detected using cross check and refined using a pixel background filling approach. Then, virtual view synthesis is performed based on disparity map and image interpolation. Finally, the noises in virtual view are removed by 5x5 median filter. In order to demonstrate the feasibility of the virtual view synthesis method, stereo image pair tsukuba from Middlebury is used.

In (Jung, Jiao, Oh, & Kim, 2010) a depth-image-based-rendering (DIBR) method is presented based on disparity map transmission over terrestrial-digital multimedia broadcasting (T-DMB). Here, left and right virtual views are created using a reference image and its corresponding depth image. Once the images have been warped, the holes are filled by bilinear interpolation. This method was evaluated subjectively using five expert viewers labeling the image sequences as bad, poor, fair, good or excellent.

A novel method to generate an accurate stereo views for an autostereoscopic 3D display is proposed in (Rhee, Choi, & Choi, 2010). Firstly, viewer's head position is estimated in real-time using a stereo camera attached on the display. Then, the synthesized view is obtained by the linear interpolation from warped left and right views. A hole filling process is performed using the disparity of the warped right and left views.

3.3.3 Discussion

According to the findings for the systematic review, image warping can be categorized in the following approaches: image domain warping, layer based and DIBR. These approaches belong to an active research field for 3DTV where goals include providing stereo to multiview conversion, low bandwidth broadcasting, among others (Fehn, 2004). Commonly found problems when applying image warping approaches for content generation include the disocclusions and holes. This problem is mainly caused because of the lack of texture information when synthesizing an image, since the reference view can have occluded regions from the virtual image point of view. In this context, the generated content is expected to be high quality, so several studies on this review are mainly focused on avoiding, detecting or smoothing holes, disoclussions and visual artifacts inherent to image warping.

The image domain warping approaches perform a mesh deformation by minimizing an energy equation. The energy equation minimization guides the mesh deformation according to estimated disparity maps, commonly sparse values given by any feature-based stereo correspondence algorithm (Yao et al., 2015). The equation can also include terms to preserve the image structure or temporal constraints for stereoscopic video. This approach avoids the disocclusions and holes that commonly appear when using DIBR at the expense of computational cost and vertical lines distortions. This warping approach implies interpolations and extrapolations inside regions of the mesh. According to (Szeliski, 1999), in a prediction error approach the error scores in prediction error not only reflect the quality of the disparity map, but also the accuracy of the selected rendering algorithm. Although this approach offers great advantages for content generation in 3DTV such as avoid disocclusions; the implicit interpolations and vertical visual artifacts that the approach commonly introduces makes it a poor prospect to perform image systhesis proposed in this thesis.

The layer based approaches simplify the DIBR problem by warping segments instead of pixels. This approaches rely in matting (Choi et al., 2013) or segmentation (Nurulfajar Abd Manap & Soraghan, 2011) techniques in order to determine the regions on the image that are going to be warped together. Once defined, the layers are warped using a single disparity value. Since, the warping process on this approach is performed by flattening the input disparity map into layers, an accuracy loss on the disparity map takes place. In the context of this thesis, using a modified version of an assessed disparity map in order to render the synthesized image does not allow its objective assessment.

The DIBR approach uses the affine disparity equation to predict the horizontal shift for each pixel in a reference image to synthesize a virtual view. DIBR techniques are commonly used to generate content for stereoscopic displays from a disparity map and a texture image. This is the case of the called shift-sensor approach. The shift-sensor approach simplifies the DIBR problem by assuming the intrinsic parameters of the two virtual stereo cameras to exactly correspond to the reference camera parameters, except for the horizontal shift of the respective principal point. The same approach can be used to synthesize intermediate views between a stereo image pairs.

This review presents the state-of-art techniques for image synthesizing. The main focus of the studies assessed on this systematic review is high quality content generation for 3DTV. Thus, several preprocessing and postprocessing techniques are used. In the context of this thesis the image synthesis processes are expected to be used as a component of the prediction error method for disparity maps quality evaluation. According to (Szeliski, 1999), error scores in prediction error not only reflect the accuracy of the disparity estimation algorithm, but also the accuracy of the selected rendering algorithm, since the rendering process of the predicted view has to deal with interpolation or extrapolation issues. Therefore DIBR techniques are shown as a good prospect to develop the evaluation method.

In this section the systematic review is defined in the context of this thesis along with the systematic reviews performed for disparity evaluation methods and image synthesis. By using the findings on the area for each field, the section 4 will present the proposed approach for disparity maps evaluation in absence of ground truth.

4. Disparity maps selection method in absence of ground truth

In a stereo vision system, a pair of rectified images of a scene is acquired from two cameras in a slightly different position. This image pair is then processed by a fixed stereo correspondence algorithm and its respective parameters to obtain a disparity map. Since the contents of the scene may vary in time according to the application field, the quality of the computed disparity maps can also vary, leading to lower quality disparity maps in some cases.

In the proposed approach the stereo vision system will compute a set of disparity maps for a scene using the same rectified image pair. Then, the quality for each computed disparity map is assessed using an evaluation method performed in absence of ground truth. The output of the proposed approach is one algorithm and its respective parameters, selected as the best candidate to build the disparity map on the assessed scene.

Figure 4.1 shows the relevant elements for the development of this proposal. In order to assess the quality of a disparity map a stereo correspondence algorithm with its respective parameters and a stereo image pair I is required. In this work the stereo image pair is assumed to be rectified, which means that the disparity search can be performed in the horizontal axis only.

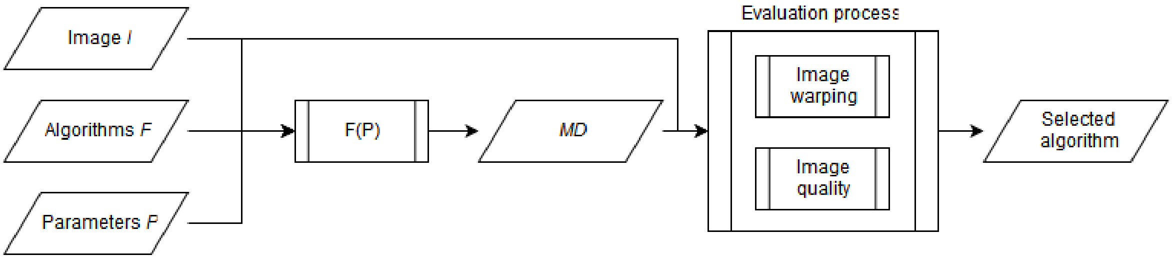


Fig. 4.1. Relevant elements on the disparity map selection process.

Here, F is the set of stereo correspondence algorithms to be assessed. P is the set of input parameters for each stereo correspondence algorithm in F. I is the stereo image pair acquired, where the stereo correspondence algorithms are going to be assessed. Through the $F(P)$ process the disparity maps MD are estimated for each stereo correspondence algorithm F_i with its respective parameters set P_i. Finally an evaluation process is performed, where a disparity map quality measure is used to select the best candidate algorithm for the scene acquired with the stereo images pair I.

The proposed evaluation method is a prediction error approach that does not require a third acquisition. Figure 4.2 shows a detailed scheme for the proposed method.

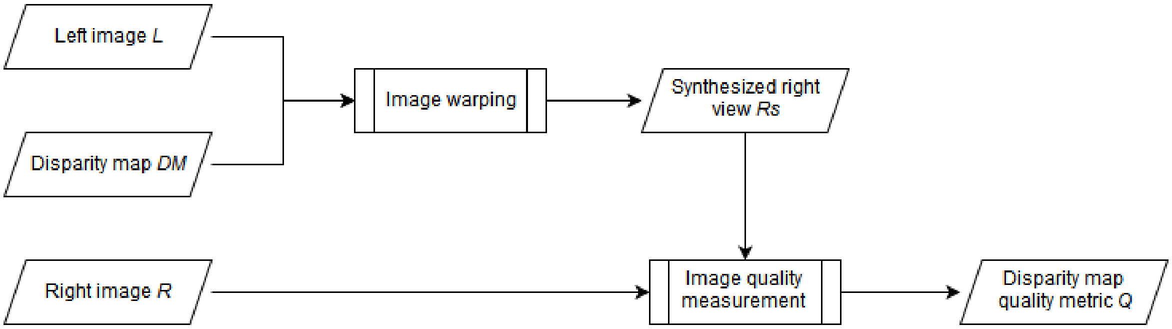

Left image *L*
Disparity map *DM*
Right image *R*
Image warping
Synthesized right view *Rs*
Image quality measurement
Disparity map quality metric *Q*

Fig. 4.2. Proposed evaluation method in absence of ground truth.

An image warping process is performed for each disparity map in MD and the input stereo image pair I. Using the DIBR image warping technique, a virtual right image R_s is synthesized from the left image and disparity map. Then, the differences between the reference right image R and the synthesized right image Rs are measured using an image quality assessment metric Q. The results obtained from the image quality measures between the reference and synthesized right views are expected to show the quality of the disparity maps for the given scene.

The following subsections present all the techniques and datasets required for developing and testing the proposed prediction error approach. The solution framework (4.1) section presents the techniques used to develop the proposed approach. The method testing (4.2) section presents the datasets required and results obtained from the developed approach testing.

4.1 Solution framework

4.1.1 Stereo correspondence algorithms

The disparity maps assessed in this approach are estimated using a set of stereo correspondence algorithms and associated parameters. As explained in section 2.2 the stereo correspondence algorithms will use a stereo image pair to compute a disparity map. The tests carried out in this work will use the opencv implementations of the block matching and the semi-global block matching algorithms (Bradski & Kaehler, 2008). The algorithms selection is made in order to ease the review and verification of the performed tests by any member of the scientific community interested in the field.

The opencv implementation of the block matching (BM) algorithm requires two mandatory parameters: *SADwindow* and *ndisp*. The SADwindow parameter defines the block size where the sum of absolute differences is computed (see section 2.2.3). The *ndisp* parameter depends of the input stereo image and defines the maximum disparity value to be found on the scene, limiting the algorithm search range. Figure 4.1.1.1 shows the Middlebury's *Piano* input stereo image pair. Figure 4.1.1.2 shows the results of computing the disparity maps for the image *Piano* from Middlebury's dataset using the BM algorithm for different SAD window sizes. The parameter *ndisp* is set to 260 according to the Middlebury's dataset documentation (Scharstein et al., 2014).

The description of the semi-global block matching (SGBM) opencv implementation is presented in section 2.2.2. In the same way that the BM algorithm the SADwindow parameter will be varied in order to obtain different quality disparity maps and the ndisp parameter will be set according to the Middlebury's documentation for the scene. Figure 4.1.1.3 shows the results of computing the disparity maps for the image *Piano* from Middlebury's dataset using the SGBM algorithm for different SAD window sizes.

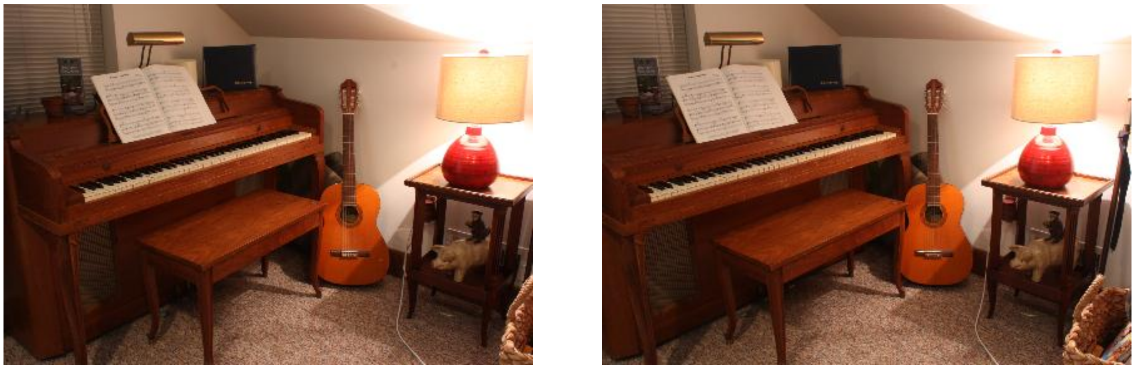
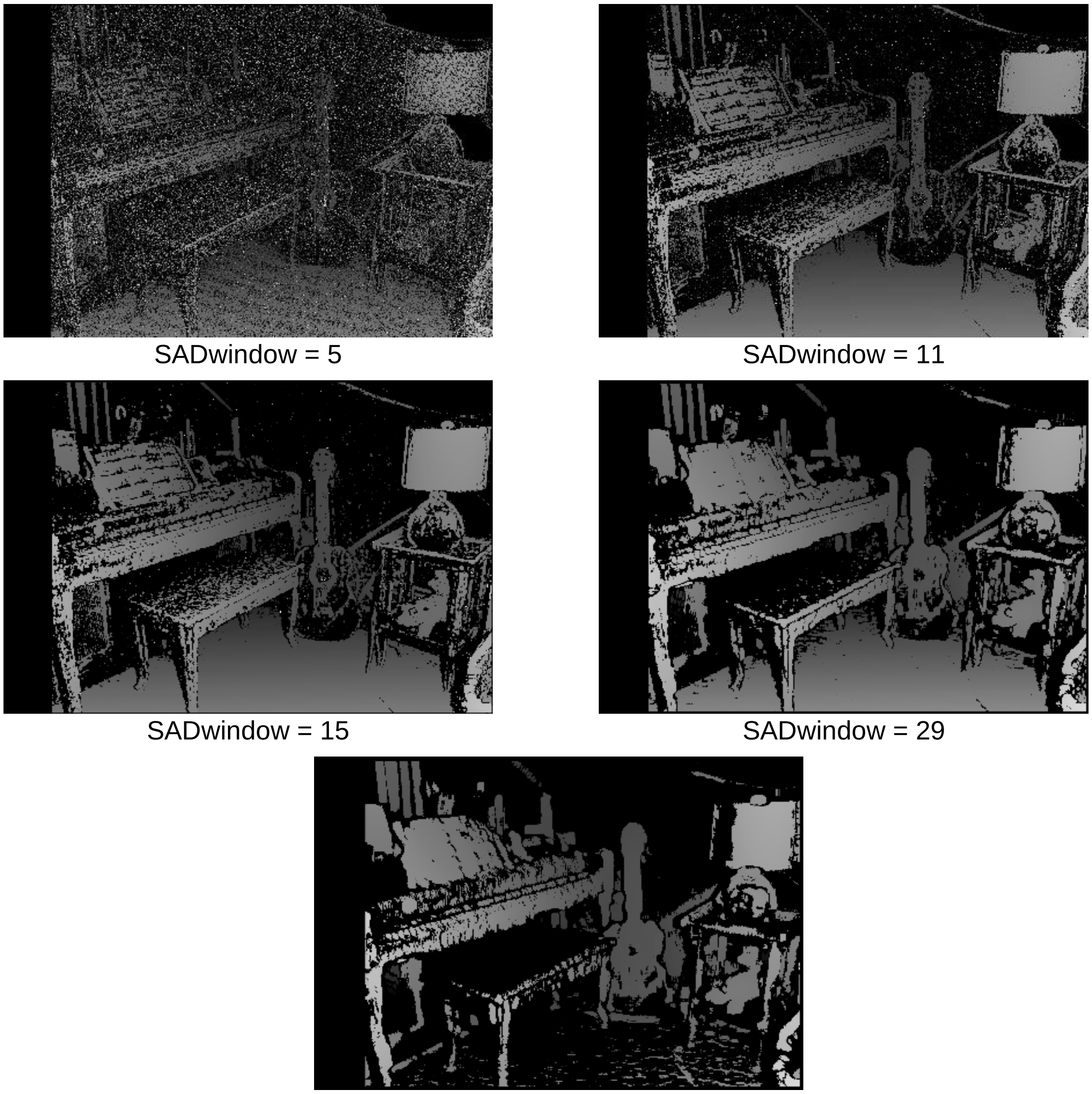


Figure 4.1.1.1. Middlebury’s *Piano* stereo image pair, left and right images respectively.



SADwindow = 5

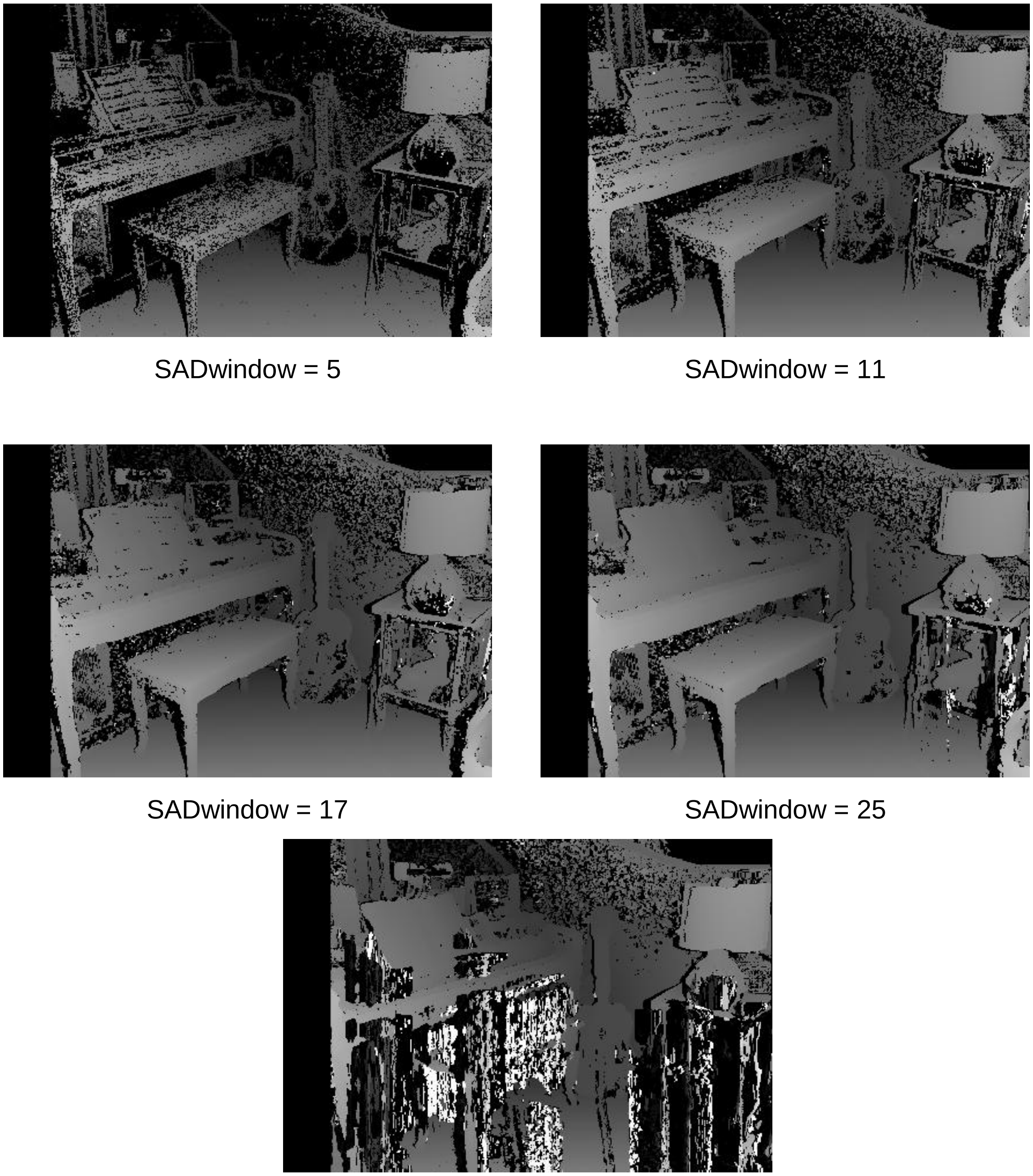
SADwindow = 11

SADwindow = 15

SADwindow = 29

SADwindow = 45

Figure 4.1.1.2. Computed disparity map for *Piano* using BM with different window sizes and n-disparities = 260.



SADwindow = 5

SADwindow = 11

SADwindow = 17

SADwindow = 25

SADwindow = 31

Figure 4.1.1.3. Computed disparity map for *Piano* using SGBM with different window sizes and *ndisp = 260*.

4.1.2 Image warping

In this work 3D image warping is used in order to synthesize virtual images to be compared with the reference acquired images. The following paragraphs briefly summarize the systematic review for image warping in the context of disparity maps evaluation.

Depth-image-based rendering (DIBR) is the process of synthesizing virtual views of a scene from still or moving color images and associated per-pixel depth information (Fehn, 2004). The DIBR technique is implemented using the affine disparity equation. The affine disparity equation can be simplified for content generation by using the shift-sensor approach. A basic warping operation can be performed using this approach, which only relies on the input stereo image pair and the respective estimated disparity map. Due its simplicity and avoidance of interpolations / extrapolations, the DIBR technique is used in the proposed approach.

In the proposed DIBR approach, the right image is synthesized using the left image and the estimated disparity map. This warping operation can be derived from the epipolar geometry in the same way than the shift-sensor approach. Figure 4.1.2.1 shows the basic stereo system geometry.

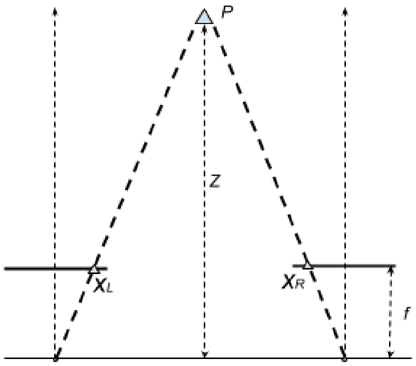


Fig. 4.1.2.1. Simplified stereo system.

Here the disparity d can be defined as:

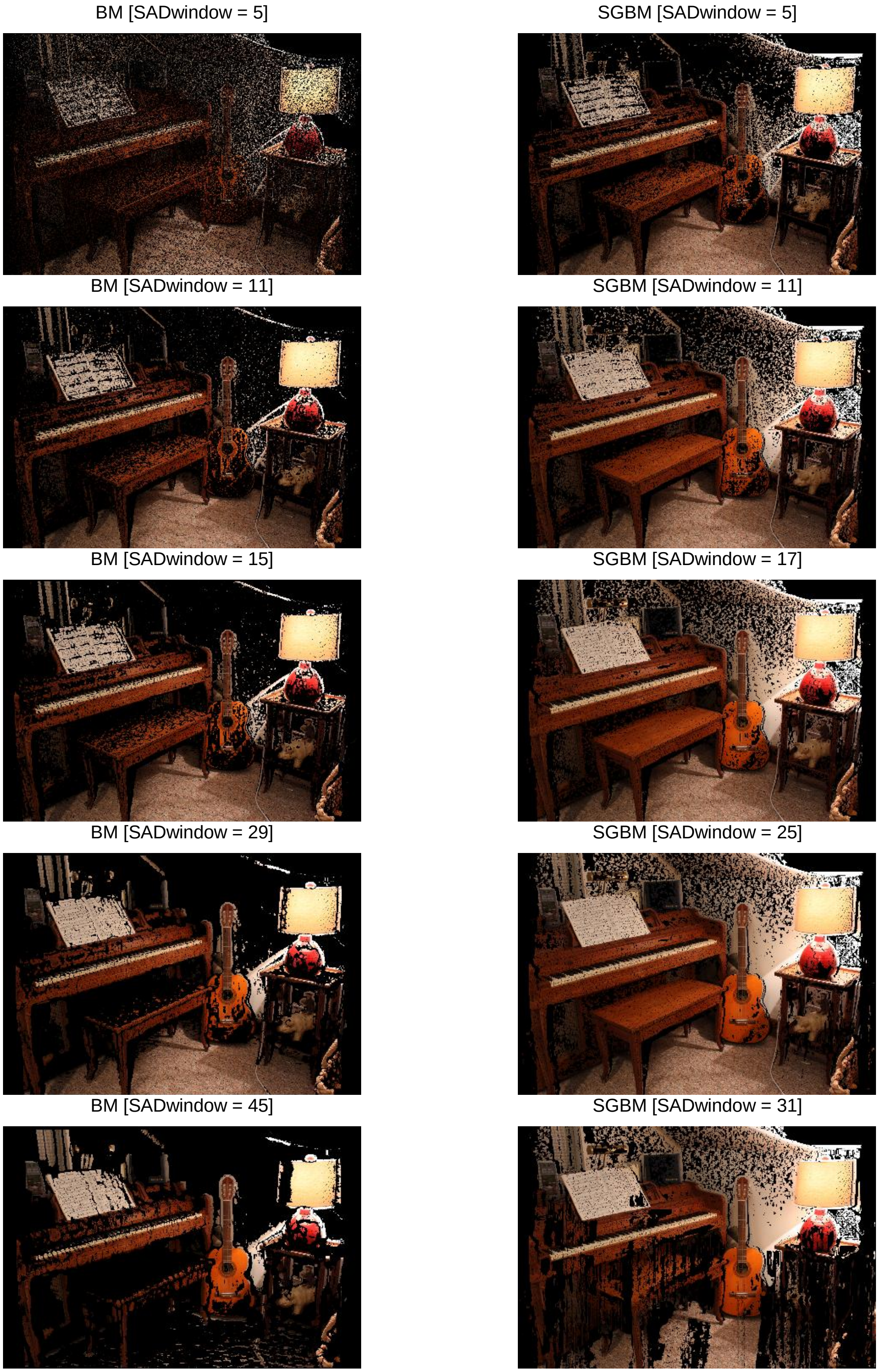
$$d = x_L - x_R \tag{18}$$

Where x_L and x_R are the horizontal pixel coordinates of the physical point P projection on the left and right images respectively. Therefore, if an estimated disparity d_{Est} is available, a virtual right image R_s can be synthesized using:

$$x_{Rs} = x_L - d_{Est} \tag{19}$$

Where x_{Rs} is the horizontal coordinate of the synthesized image R_s corresponding to x_L pixel in the left image. Table 4.1.2.1 shows the warped right image for Middlebury's *Piano* using each one of the disparity maps presented on tables 4.1.1.1 and 4.1.1.2.

From table 4.1.2.1 can be noticed that different image synthesizing results are obtained from different disparity maps for the same scene. Additionally, disocclussions are an inherent property on the proposed 3D warping method, since in no left image will contain all the information to reproduce the right image. Hence, no synthesized image is expected to obtain an ideal score on the image quality assessment step.



BM [SADwindow = 5]

SGBM [SADwindow = 5]

BM [SADwindow = 11]

SGBM [SADwindow = 11]

BM [SADwindow = 15]

SGBM [SADwindow = 17]

BM [SADwindow = 29]

SGBM [SADwindow = 25]

BM [SADwindow = 45]

SGBM [SADwindow = 31]

Table 4.1.2.1. *Piano* synthesized right image using disparity maps from section 4.1.1.

4.1.3 Image quality assessment

Once the synthesized right image has been estimated, a comparison between the right reference image and right synthesized image is required. Since an objective evaluation is ideal and a reference image is available, a full-reference image quality assessment is performed. Table 4.1.3.1 shows the computed image quality metrics on Middlebury's *Piano* measuring the quality of the synthesized right images from section 4.1.2 compared against the reference right image. For this assessment the MSSIM, PSNR, MSE, PMSE, MAE and AD metrics are used (section 2.3).

Algorithm	MSSIM	PSNR	MSE	PMSE	MAE	AD
BM [SADwindow = 5]	0.16	8	10295.08	40.37	66.22	63.9
BM [SADwindow = 11]	0.26	8.39	9417	36.93	59.55	58.48
BM [SADwindow = 15]	0.29	8.54	9105.77	35.71	57.5	56.49
BM [SADwindow = 29]	0.3	8.65	8863.49	34.76	56.32	55.34
BM [SADwindow = 45]	0.18	8.03	10243.15	40.17	67.88	67.28
SGBM [SADwindow = 7]	0.27	9.07	8054.8	31.59	52.36	51.46
SGBM [SADwindow = 9]	0.4	10.21	6195.73	24.3	40.07	38.89
SGBM [SADwindow = 11]	0.46	10.96	5218.83	20.47	33.71	32.38
SGBM [SADwindow = 15]	0.5	11.44	4666.95	18.3	30.22	28.75
SGBM [SADwindow = 17]	0.31	10.59	5675.08	22.26	40.33	37.94

Table 4.1.3.1. Image quality metrics computed for *Piano* reference and synthesized right images using algorithms presented in section 4.2.2.

In recent years the structural similarity index (SSIM) has become an accepted standard among image quality metrics. Made up of three components, this technique assesses the visual impact of changes in image luminance, contrast, and structure locally (Dosselmann & Yang, 2011). Given the nature of the assessed images in the context of this thesis, where the synthesized images are expected to show local structural variations depending of the disparity map quality, the MSSIM metric is selected as the image quality measurement to be used in the automated disparity map selection method.

However, (Dosselmann & Yang, 2011) proves empirically and formally that the MSSIM can perform poorly when computing the quality of visually different corrupted images by assigning similar scores, in the same way that MSE does. For this reason, is expected that some ties in MSSIM scores will show up when assessing visually different synthesized images. The proposed tie-breaker strategy in this thesis consists in assessing the MSE scores when the MSSIM scores are tied for the winning algorithms.

In accordance with the stated techniques and selection strategies proposed in this section, the disparity map computed using SGBM and *SADwindow = 15* is selected. Figure 4.1.3.1 shows the disparity map and warped image for selected candidate. The selected disparity map is presumably the one with lower error rate among the assessed set.

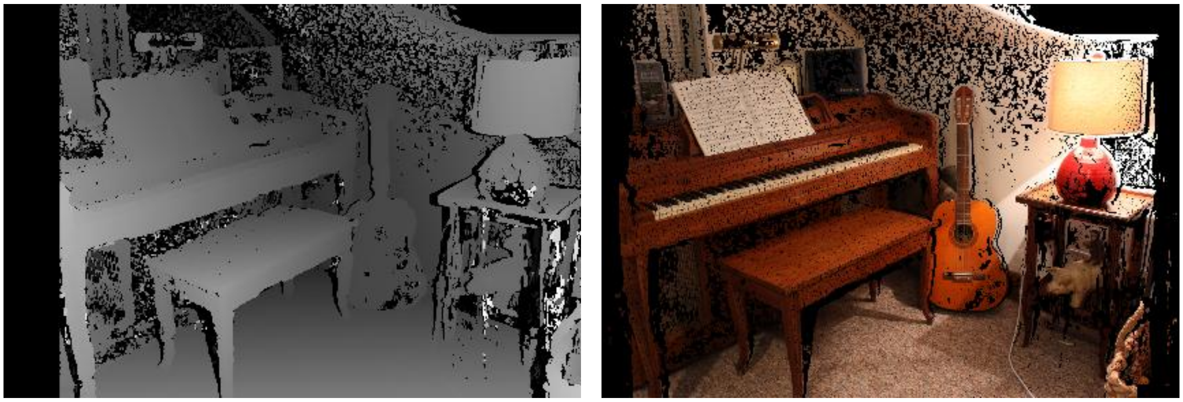


Fig. 4.1.3.1. Disparity map and synthesized image for the selected stereo correspondence algorithm (SGBM [*SADwindow = 15*])

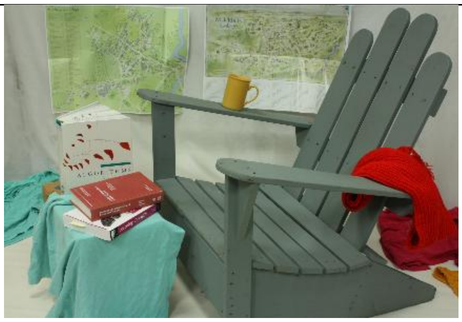
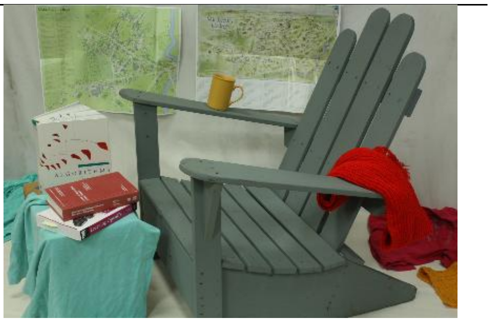
4.2 Proposed method Testing

This section presents two approaches of testing for the proposed method. Firstly, The Middlebury dataset on its third version used for testing purposes is presented on section 4.2.1. Secondly, the results of the proposed approach on each step are presented in section 4.2.2. Finally, a comparison between the proposed selection method and the popular error measuring approach using ground truth data is performed in section 4.2.3. The datasets, stereo correspondence algorithms and parameters used in this section where chosen for two main reasons: to generate disparity maps of different quality for a particular scene and to guarantee the availability of the information in order to validate or verify any of the conducted tests.

4.2.1 Datasets

This work uses the Middlebury's training dataset from its third version (Scharstein et al., 2014) in order to test the proposed approach. The Middlebury's dataset is a standard for general purpose stereo correspondence algorithms evaluation. Table 4.2.1.1 shows the stereo pair images available on the dataset along with their parameter *ndisp* that stereo correspondence algorithms require to compute the disparity maps.

This dataset available at Middlebury's website is composed of ten stereo image pairs: Adirondack, Jadeplant, Motorcycle, Piano, Pipes, Playroom, Playtable, Recycle, Shelves, and Vintage, along with information about the camera calibration and maximum disparity for each scene.

Image name	**Left image**	**Right image**
Adirondack [ndisp=280]		

Jadeplant

[ndisp=640]

Motorcycle

[ndisp=270]

Piano

[ndisp=260]

Pipes

[ndisp=300]

Playroom

[ndisp=330]

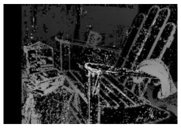
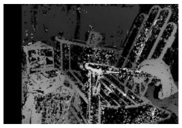
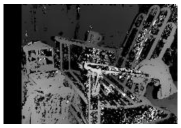
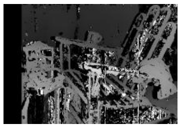
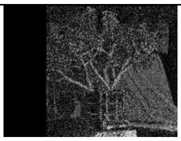
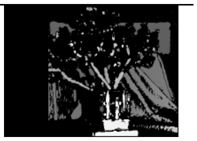
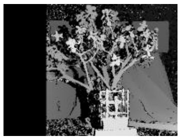
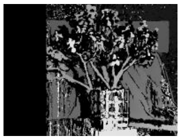
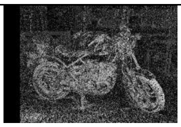
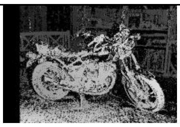
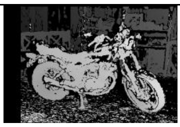
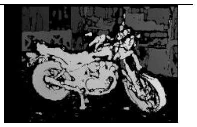
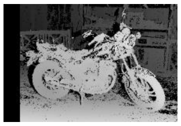
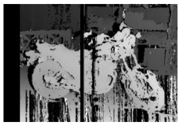
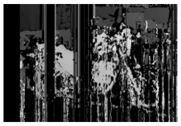
Playtable [ndisp=290]		
Recycle [ndisp=260]		
Shelves [ndisp=240]		
Vintage [ndisp=740]		

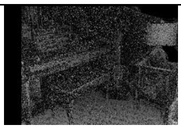
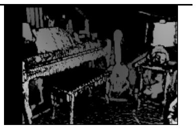
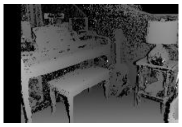
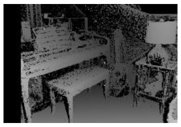
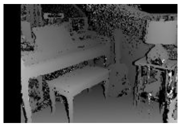
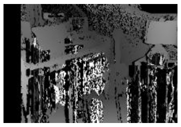
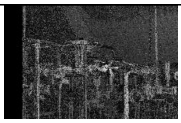
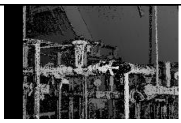
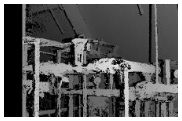
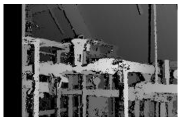
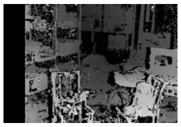
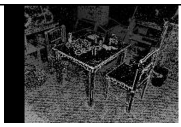
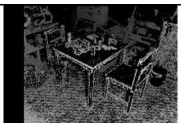
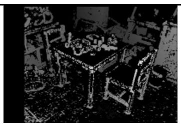
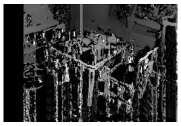
Table 4.1.1.1. Middlebury's version 3 training dataset.

4.2.2 Testing on Middlebury's dataset

The following section presents the results of testing the proposed method on the stereo image dataset (4.2.1). For each stereo image ten disparity maps will be estimated using the algorithms and parameters presented in section 4.1.1. Then, a 3D image warping process will be performed as stated in section 4.1.2. Finally, the SSIM metric is computed for each synthesized right image and the best score is selected.

Table 4.2.2.1 shows the results on estimating disparity maps for each image on the testing dataset by using the algorithms BM [SADwindow = 5, 11, 15, 29, 45] and SGBM [SADwindow = 7, 9, 11, 15, 17] as explained in section 4.1.1. The parameter selection to compute the disparity maps is made arbitrary in order to guarantee disparity maps with different quality levels.

Image	**Disparity maps**				
Adirondack	 a.	 b.	 c.	 d.	 e.
	 f.	 g.	 h.	 i.	 j.
Jadeplant	 a.	 b.	 c.	 d.	 e.
	 f.	 g.	 h.	 i.	 j.
Motorcycle	 a.	 b.	 c.	 d.	 e.
	 f.	 g.	 h.	 i.	 j.

Piano					
	a.	b.	c.	d.	e.
					
	f.	g.	h.	i.	j.
Pipes					
	a.	b.	c.	d.	e.
					
	f.	g.	h.	i.	j.
Playroom					
	a.	b.	c.	d.	e.
					
	f.	g.	h.	i.	j.
Playtable					
	a.	b.	c.	d.	e.
					
	d.	e.	f.	g.	h.

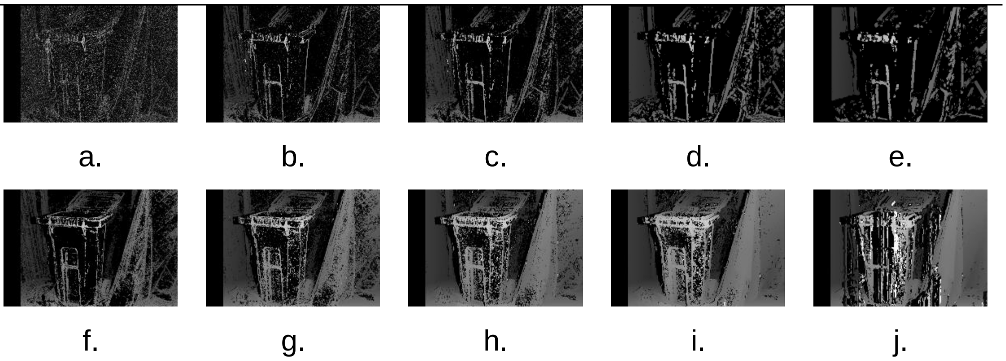
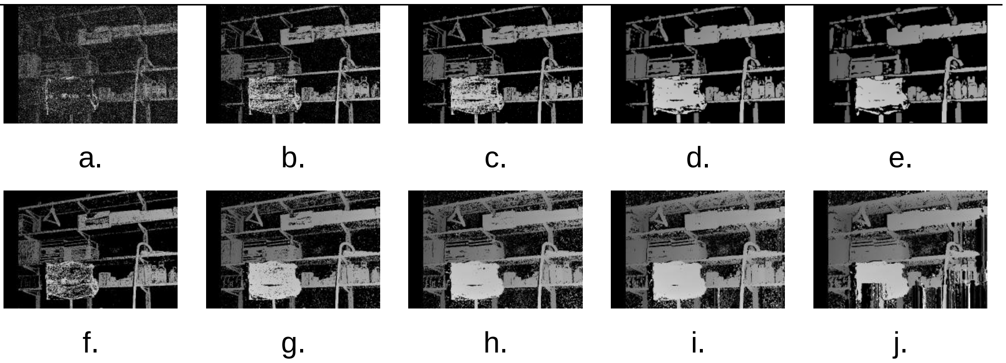
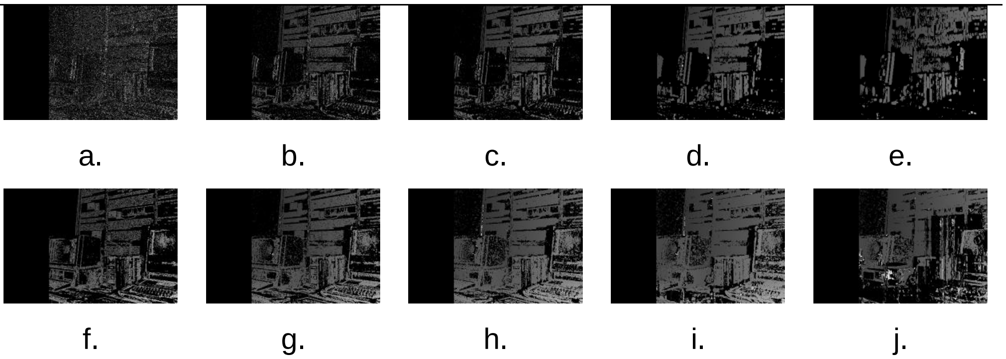
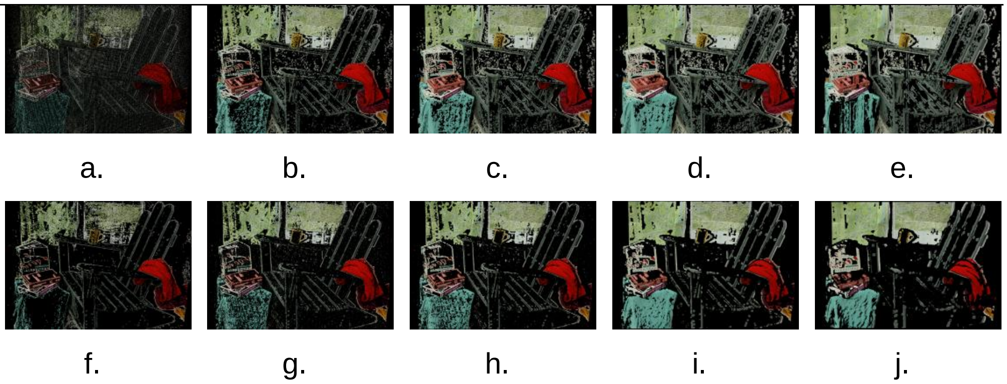
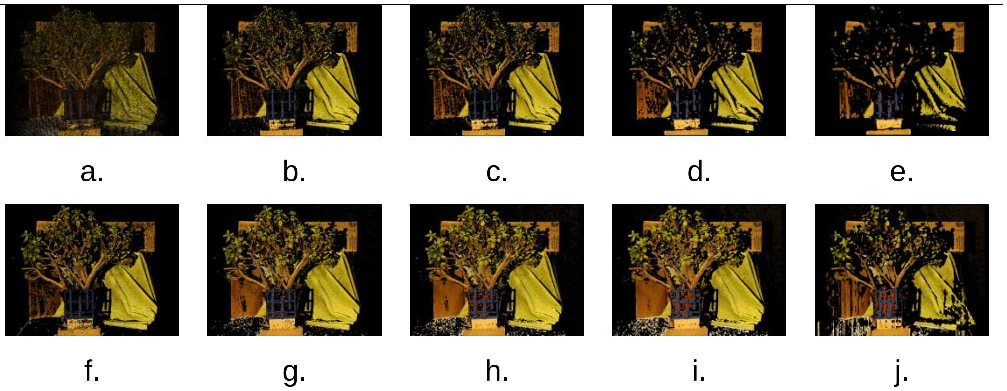
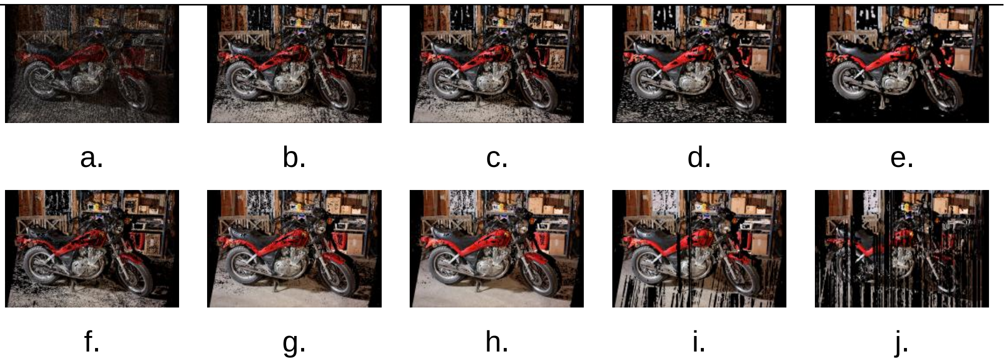
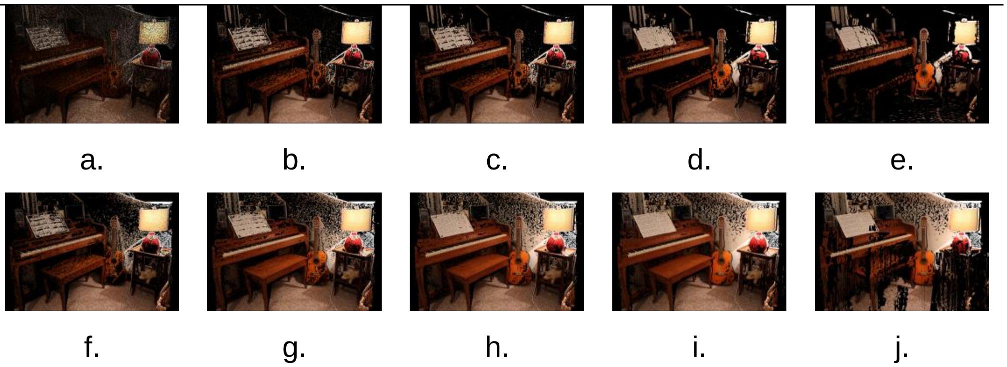
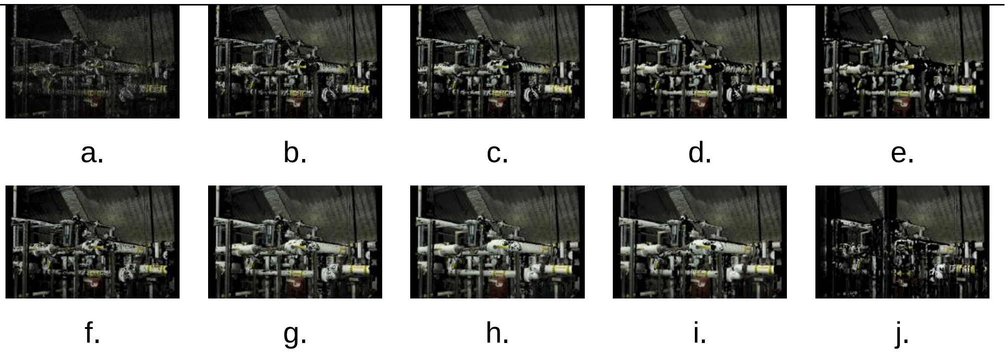
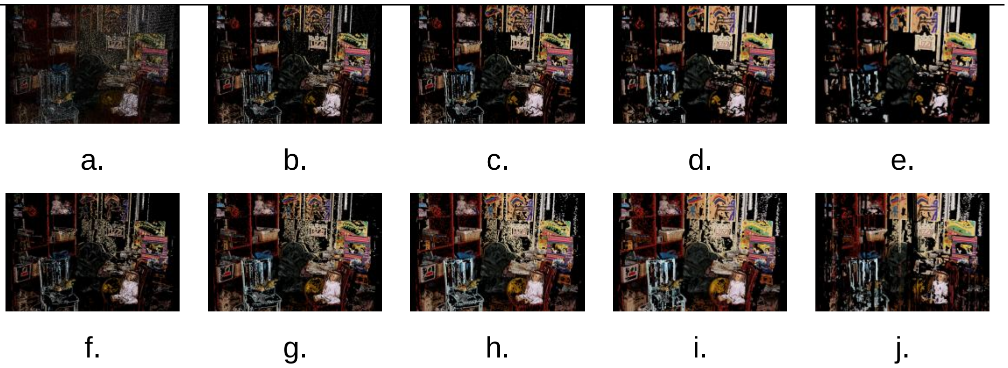
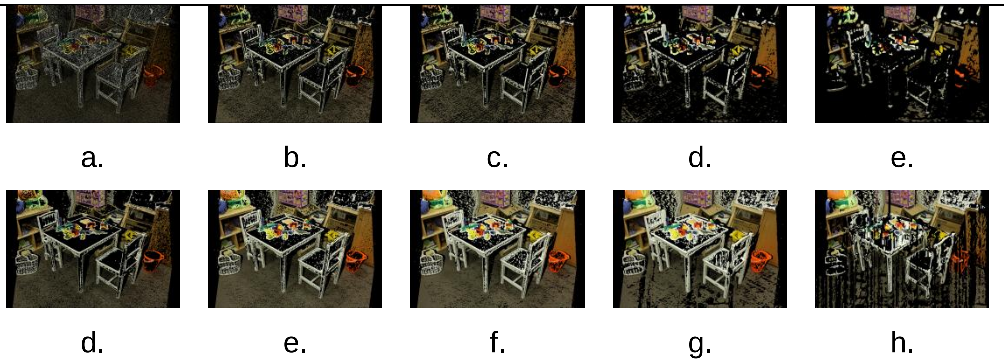
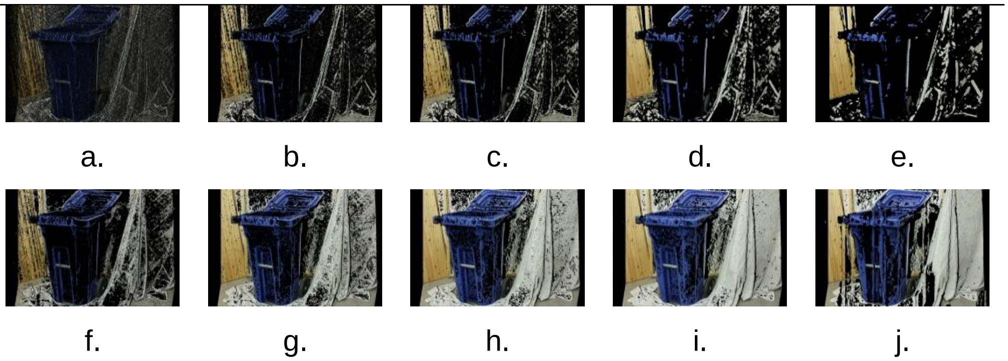
Recycle	
Shelves	
Vintage	

Table 4.2.2.1. Disparity maps computed for Middlebury's dataset using the stereo correspondence algorithms in presented section 4.1.1. (a) to (e) BM with SADwindows 5, 11, 15, 29, 45, respectively. (f) to (j) SGBM with SADwindows 7, 9, 11, 15, 17, respectively.

Table 4.2.2.2 shows the respective synthesized right image for each disparity map using only the left image. The completeness and accuracy of the synthesized image is derived from the disparity map quality. Visual differences for each synthesized image can be noticed according to each disparity map in table 4.2.2.1.

Image	**Synthesized Images**
Adirondack	 a. b. c. d. e. f. g. h. i. j.
Jadeplant	 a. b. c. d. e. f. g. h. i. j.
Motorcycle	 a. b. c. d. e. f. g. h. i. j.
Piano	 a. b. c. d. e. f. g. h. i. j.

Pipes	
Playroom	
Playtable	
Recycle	

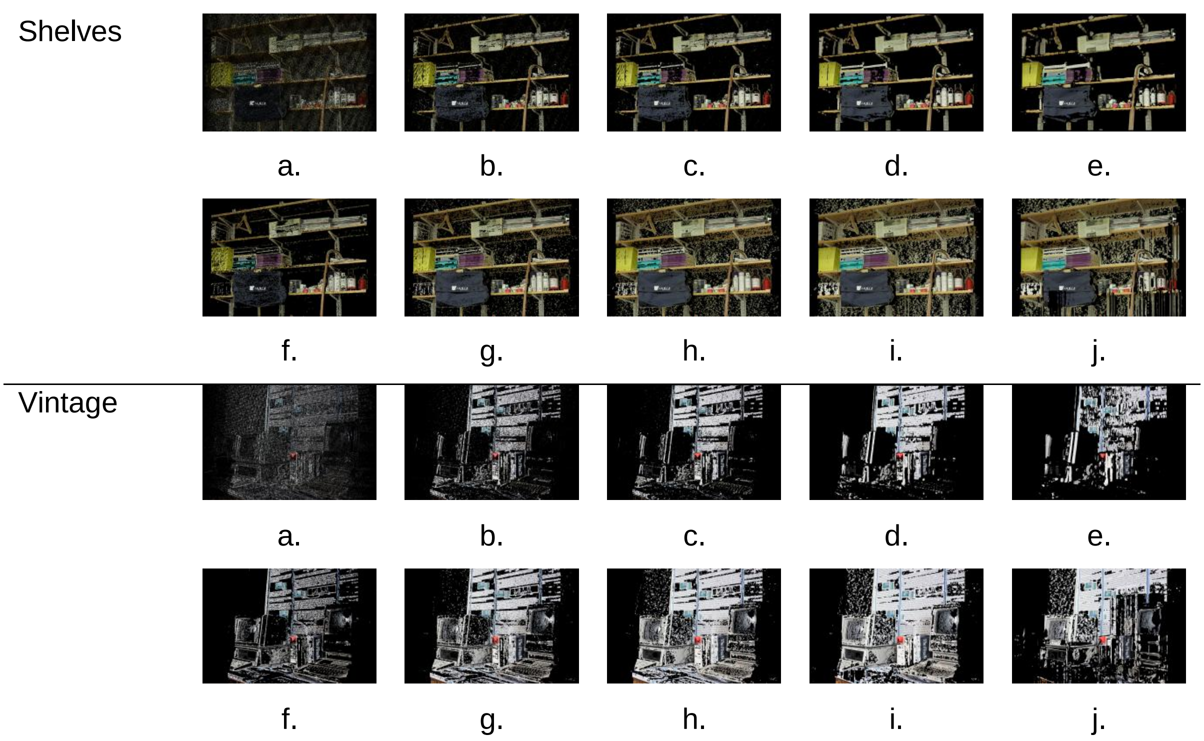


Table 4.2.2.2. Syntesized images computed for Middlebury's dataset and the stereo correspondence algorithms in table 4.2.2.1, respectively.

Tables 4.2.2.3 to 4.2.2.12 shows the computed 2D image quality metrics presented in section 2.3 for each right image in the dataset (4.2.1) and each warped image in table 4.2.2.2.

Algorithm	MSSIM	PSNR	MSE	PMSE	MAE	AD
BM [SADwindow = 5]	**0.05**	7.85	10657.92	45.74	84.23	79.36
BM [SADwindow = 11]	**0.11**	8.17	9920.71	42.58	79.02	74.47
BM [SADwindow = 15]	**0.14**	8.39	9412.97	40.4	75.48	70.53
BM [SADwindow = 29]	**0.2**	8.74	8700.47	37.18	70.73	65.23
BM [SADwindow = 45]	**0.21**	8.63	8913.17	38.09	72.57	67.33
SGBM [SADwindow = 7]	**0.06**	7.63	11230.92	48.2	87.59	84.98
SGBM [SADwindow = 9]	**0.16**	8.82	8531.48	36.77	68.4	63.44
SGBM [SADwindow = 11]	**0.23**	9.66	7026.17	30.29	57.44	51.08
SGBM [SADwindow = 15]	**0.29**	10.35	5996.69	25.85	49.91	42.5
SGBM [SADwindow = 17]	**0.30**	10.21	6200.88	26.73	51.25	43.98

Table 4.2.2.3. Image quality metrics computed for ***Adirondack*** reference and synthesized right images.

Algorithm	MSSIM	PSNR	MSE	PMSE	MAE	AD
BM [SADwindow = 5]	**0.09**	8.43	9328.26	36.58	63.54	59.99
BM [SADwindow = 11]	**0.13**	8.91	8366.49	32.81	56.92	54.87
BM [SADwindow = 15]	**0.14**	8.94	8300.71	32.55	56.33	54.52
BM [SADwindow = 29]	**0.14**	8.67	8823.02	34.6	59.4	57.94
BM [SADwindow = 45]	**0.11**	8.2	9849.93	38.63	65.82	64.67
SGBM [SADwindow = 7]	**0.13**	9.14	7926.23	31.08	53.92	52.5
SGBM [SADwindow = 9]	**0.23**	9.92	6619.61	25.96	45.49	43.37
SGBM [SADwindow = 11]	**0.28**	10.36	5987.93	23.48	41.6	39.01
SGBM [SADwindow = 15]	**0.31**	10.59	5672.17	22.24	39.68	36.77
SGBM [SADwindow = 17]	**0.24**	9.46	7366.7	28.89	50.05	48.06

Table 4.2.2.4. Image quality metrics computed for ***Jadeplant*** reference and synthesized right images.

Algorithm	MSSIM	PSNR	MSE	PMSE	MAE	AD
BM [SADwindow = 5]	**0.14**	8.1	10063.74	39.47	72.95	68.03
BM [SADwindow = 11]	**0.26**	9.58	7170.35	28.12	53.22	49.47
BM [SADwindow = 15]	**0.31**	10.08	6377.03	25.01	48.02	44.18
BM [SADwindow = 29]	**0.35**	9.21	7802.91	30.6	56.42	52.91
BM [SADwindow = 45]	**0.34**	8.59	8997.05	35.28	64.13	60.88
SGBM [SADwindow = 7]	**0.28**	9.68	7000.5	27.45	51.13	48.26
SGBM [SADwindow = 9]	**0.43**	12.23	3895.4	15.28	31.36	27.17
SGBM [SADwindow = 11]	**0.52**	13.33	3019.88	11.84	25.85	21.05
SGBM [SADwindow = 15]	**0.45**	11.32	4793.72	18.8	37.4	32.86
SGBM [SADwindow = 17]	**0.21**	8.59	9004.18	35.31	66.67	60.84

Table 4.2.2.5. Image quality metrics computed for ***Motorcycle*** reference and synthesized right images.

Algorithm	MSSIM	PSNR	MSE	PMSE	MAE	AD
BM [SADwindow = 5]	**0.16**	8	10295.08	40.37	66.22	63.9
BM [SADwindow = 11]	**0.26**	8.39	9417	36.93	59.55	58.48
BM [SADwindow = 15]	**0.29**	8.54	9105.77	35.71	57.5	56.49
BM [SADwindow = 29]	**0.3**	8.65	8863.49	34.76	56.32	55.34
BM [SADwindow = 45]	**0.18**	8.03	10243.15	40.17	67.88	67.28
SGBM [SADwindow = 7]	**0.27**	9.07	8054.8	31.59	52.36	51.46
SGBM [SADwindow = 9]	**0.4**	10.21	6195.73	24.3	40.07	38.89
SGBM [SADwindow = 11]	**0.46**	10.96	5218.83	20.47	33.71	32.38
SGBM [SADwindow = 15]	**0.5**	11.44	4666.95	18.3	30.22	28.75
SGBM [SADwindow = 17]	**0.31**	10.59	5675.08	22.26	40.33	37.94

Table 4.2.2.6. Image quality metrics computed for ***Piano*** reference and synthesized right images.

Algorithm	MSSIM	PSNR	MSE	PMSE	MAE	AD
BM [SADwindow = 5]	**0.17**	9.99	6517.14	25.56	52.88	48.32
BM [SADwindow = 11]	**0.37**	10.93	5249.28	20.59	41.11	37.24
BM [SADwindow = 15]	**0.41**	11.2	4934.65	19.35	38.73	34.79
BM [SADwindow = 29]	**0.45**	11.24	4882.82	19.15	37.86	33.74
BM [SADwindow = 45]	**0.44**	10.81	5392.87	21.15	41.07	36.95
SGBM [SADwindow = 7]	**0.38**	11.33	4790.66	18.79	37.9	34.72
SGBM [SADwindow = 9]	**0.51**	12.63	3550.94	13.93	28.93	24.69
SGBM [SADwindow = 11]	**0.57**	13.09	3194.1	12.53	26.26	21.51
SGBM [SADwindow = 15]	**0.56**	12.83	3385.58	13.28	27.57	22.39
SGBM [SADwindow = 17]	**0.35**	10.44	5877.77	23.05	46.37	41.95

Table 4.2.2.7. Image quality metrics computed for ***Pipes*** reference and synthesized right images.

Algorithm	MSSIM	PSNR	MSE	PMSE	MAE	AD
BM [SADwindow = 5]	**0.13**	7.59	11314.05	44.37	77.1	69.7
BM [SADwindow = 11]	**0.17**	7.76	10880.4	42.67	73.15	67.46
BM [SADwindow = 15]	**0.19**	7.89	10573.38	41.46	71.2	65.36
BM [SADwindow = 29]	**0.22**	8.12	10015.3	39.28	68.47	62.44
BM [SADwindow = 45]	**0.22**	8.13	10004.67	39.23	69.19	63.59
SGBM [SADwindow = 7]	**0.15**	7.61	11280.92	44.24	75.16	71.53
SGBM [SADwindow = 9]	**0.25**	8.53	9124.96	35.78	61.46	54.94
SGBM [SADwindow = 11]	**0.32**	9.21	7808.01	30.62	53.51	45.39
SGBM [SADwindow = 15]	**0.36**	9.71	6948.24	27.25	48.63	39.54
SGBM [SADwindow = 17]	**0.23**	8.76	8655.27	33.94	61.43	52.78

Table 4.2.2.8. Image quality metrics computed for ***Playroom*** reference and synthesized right images.

Algorithm	MSSIM	PSNR	MSE	PMSE	MAE	AD
BM [SADwindow = 5]	**0.09**	7.03	12871.06	50.47	82.84	80.06
BM [SADwindow = 11]	**0.13**	7.05	12818.32	50.27	78.64	76.65
BM [SADwindow = 15]	**0.14**	7.13	12600.19	49.41	77.43	75.49
BM [SADwindow = 29]	**0.13**	6.89	13297.91	52.15	85.88	84.63
BM [SADwindow = 45]	**0.1**	6.5	14540.94	57.02	93.6	92.58
SGBM [SADwindow = 7]	**0.15**	7.37	11922.97	46.76	72.51	70.67
SGBM [SADwindow = 9]	**0.28**	8.4	9403.93	36.88	56.06	53.49
SGBM [SADwindow = 11]	**0.35**	9.21	7800.28	30.59	47.6	44.66
SGBM [SADwindow = 15]	**0.34**	9.71	6950.71	27.26	45.42	42.45
SGBM [SADwindow = 17]	**0.18**	8.39	9422.08	36.95	63.74	60.32

Table 4.2.2.9. Image quality metrics computed for ***Playtable*** reference and synthesized right images.

Algorithm	MSSIM	PSNR	MSE	PMSE	MAE	AD
BM [SADwindow = 5]	**0.02**	5.46	18477.92	72.46	113.1	110.3
BM [SADwindow = 11]	**0.04**	5.42	18676.75	73.24	112.89	111.46
BM [SADwindow = 15]	**0.05**	5.58	17972.49	70.48	108.68	107.29
BM [SADwindow = 29]	**0.07**	5.69	17550.79	68.83	106.63	105.28
BM [SADwindow = 45]	**0.07**	5.43	18604.86	72.96	113.16	112.06
SGBM [SADwindow = 7]	**0.05**	6.1	15975.94	62.65	97.56	96.31
SGBM [SADwindow = 9]	**0.12**	7.88	10598.45	41.56	66.66	64.47
SGBM [SADwindow = 11]	**0.18**	9.08	8039.65	31.53	51.8	49.09
SGBM [SADwindow = 15]	**0.25**	9.92	6627.02	26.09	43.49	40.41
SGBM [SADwindow = 17]	**0.21**	8.83	8519.36	33.54	55.31	52.17

Table 4.2.2.10. Image quality metrics computed for ***Recycle*** reference and synthesized right images.

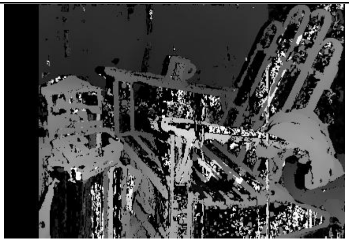
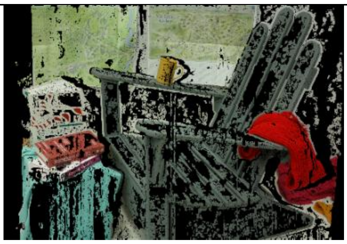
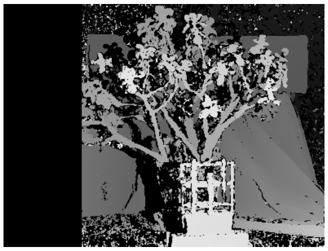
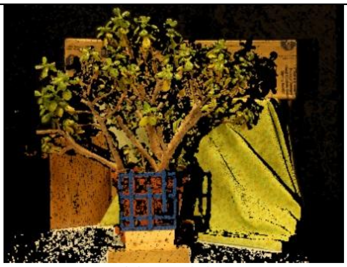
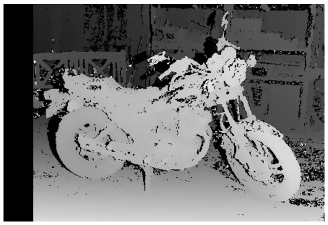
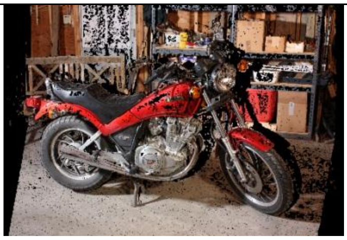
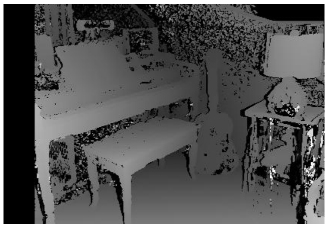
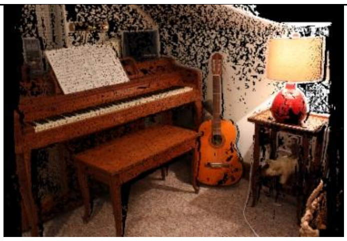
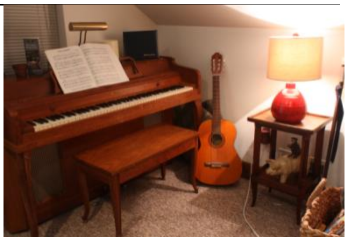
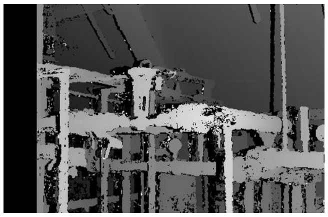
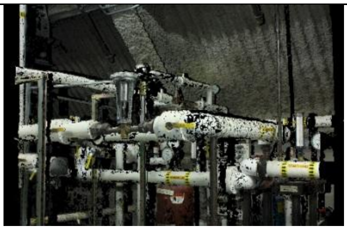
Algorithm	MSSIM	PSNR	MSE	PMSE	MAE	AD
BM [SADwindow = 5]	**0.06**	7.69	11072.4	43.42	85.33	83.05
BM [SADwindow = 11]	**0.1**	7.72	11001.76	43.14	84.86	82.97
BM [SADwindow = 15]	**0.14**	7.82	10739.69	42.12	82.75	80.72
BM [SADwindow = 29]	**0.21**	8.03	10223.64	40.09	79	76.68
BM [SADwindow = 45]	**0.23**	8.1	10074.85	39.51	78.24	75.87
SGBM [SADwindow = 7]	**0.1**	7.72	10993.05	43.11	84.44	82.8
SGBM [SADwindow = 9]	**0.2**	8.59	8987.22	35.24	68.49	65.99
SGBM [SADwindow = 11]	**0.28**	9.37	7523.24	29.5	57.26	54.3
SGBM [SADwindow = 15]	**0.33**	10.09	6375.13	25.2	48.71	45.37
SGBM [SADwindow = 17]	**0.31**	9.97	6545.09	25.67	50.07	46.54

Table 4.2.2.11. Image quality metrics computed for ***Shelves*** reference and synthesized right images.

Algorithm	MSSIM	PSNR	MSE	PMSE	MAE	AD
BM [SADwindow = 5]	**0.01**	2.86	33679.97	132.08	163.45	161.05
BM [SADwindow = 11]	**0.03**	2.88	33527.84	131.48	161.82	160.67
BM [SADwindow = 15]	**0.06**	3.01	32484.15	127.39	157.15	155.97
BM [SADwindow = 29]	**0.11**	3.2	31148.94	122.15	151.32	150.15
BM [SADwindow = 45]	**0.06**	2.96	32902.66	129.03	159.29	158.34
SGBM [SADwindow = 7]	**0.02**	2.88	33510.19	131.41	159.9	158.91
SGBM [SADwindow = 9]	**0.1**	3.69	27782.57	108.95	133.82	132.28
SGBM [SADwindow = 11]	**0.17**	4.3	24149.33	94.7	117.44	115.65
SGBM [SADwindow = 15]	**0.21**	4.61	22482.13	88.17	109.77	107.86
SGBM [SADwindow = 17]	**0.14**	3.71	27667	108.5	134	132.16

Table 4.2.2.12. Image quality metrics computed for ***Vintage*** reference and synthesized right images.

The table 4.2.2.13 shows the algorithm selected for each scene with its corresponding disparity map, synthesized and reference right images.

Image	Disparity map	Synthesized image	Right image
Adirondack	 SGBM [*SADwindow = 17*]	 MSSIM = 0.30	
Jadeplant	 SGBM [*SADwindow = 15*]	 MSSIM = 0.31	
Motorcycle	 SGBM [*SADwindow = 11*]	 MSSIM = 0.52	
Piano	 SGBM [*SADwindow = 15*]	 MSSIM = 0.50	
Pipes	 SGBM [*SADwindow = 11*]	 MSSIM = 0.57	

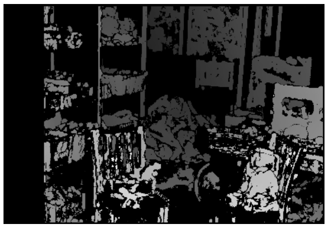
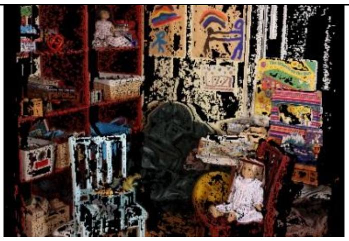
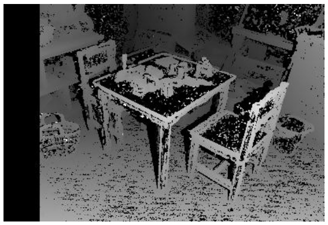
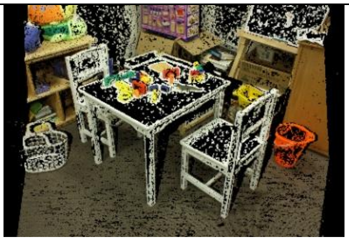
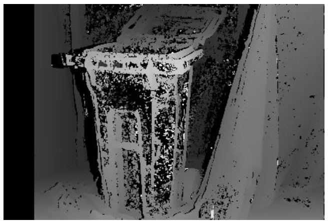
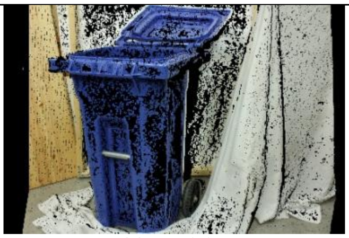
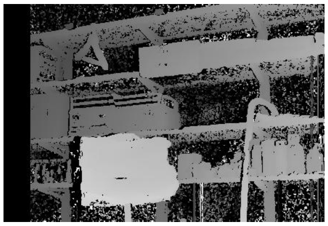
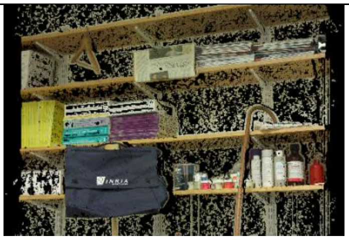
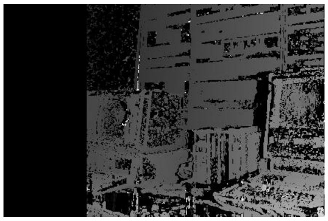
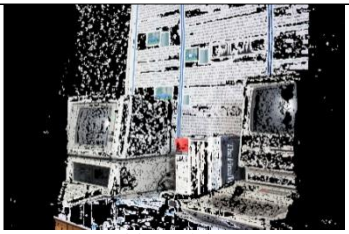
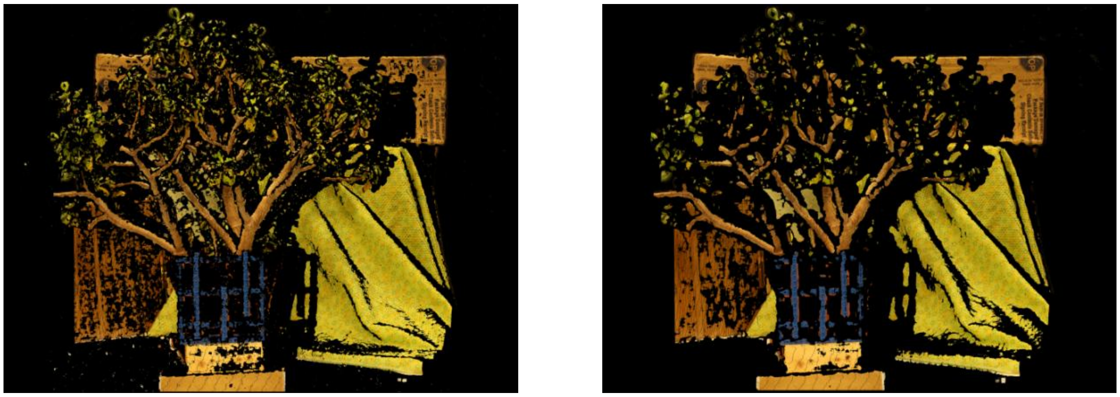
Playroom	 SGBM [*SADwindow = 15*]	 MSSIM = 0.36	
Playtable	 SGBM [*SADwindow = 11*]	 MSSIM = 0.35	
Recycle	 SGBM [*SADwindow = 15*]	 MSSIM = 0.25	
Shelves	 SGBM [*SADwindow = 15*]	 MSSIM = 0.33	
Vintage	 SGBM [*SADwindow = 15*]	 MSSIM = 0.21	

Table 4.2.2.13. Selected stereo correspondence algorithms.

From table 4.2.2.13, the algorithm SGBM with a *SADwindow* = *15* parameter is consistently shown as the best candidate according to the proposed method, only surpassed by the same algorithm with a *SADwindow* = *11* in Motorcycle, Pipes, Playtable scenes and *SADwindow* = *17* in the *Adirondack* scene.

The particular cases presented on this section did not involve any MSSIM scores tie for the selected disparity map candidates. Therefore, the candidate disparity maps were selected by using the MSSIM only. Nevertheless, this is a feasible case scenario in which according to the proposal the MSE computation will be required and used as a tie-breaker. Figure 4.2.2.1 shows a case of tie scores for *Jadeplant* right synthesized images using the BM stereo algorithm with *SADwindow* of 15 and 29.



BM [*SADwindow = 15*]
MSSIM = 0.14, MSE = 8300.71

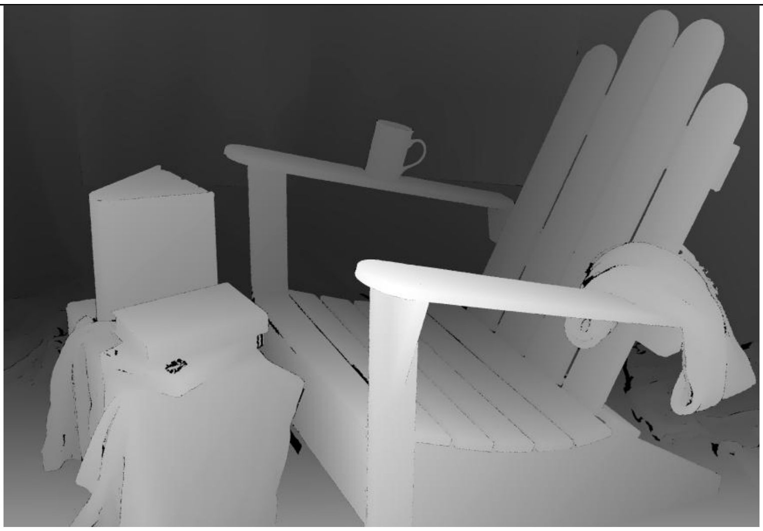
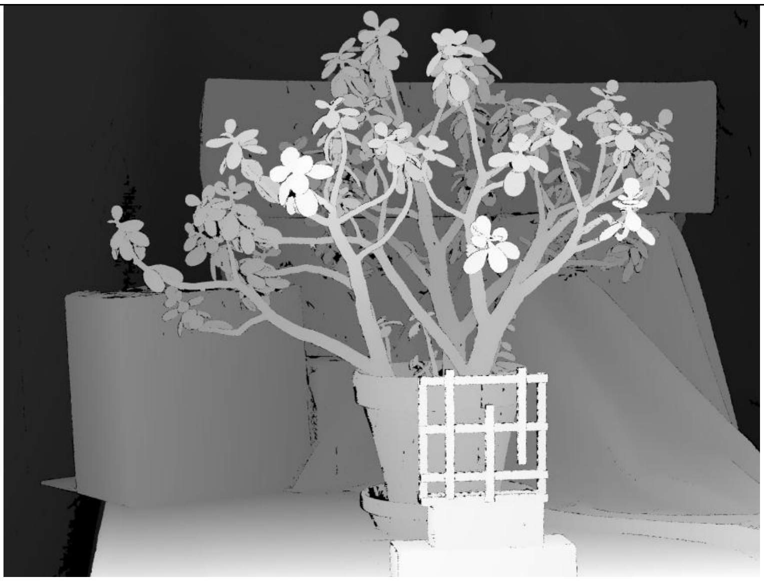
BM [*SADwindow = 29*]
MSSIM = 0.14, MSE = 8823.02

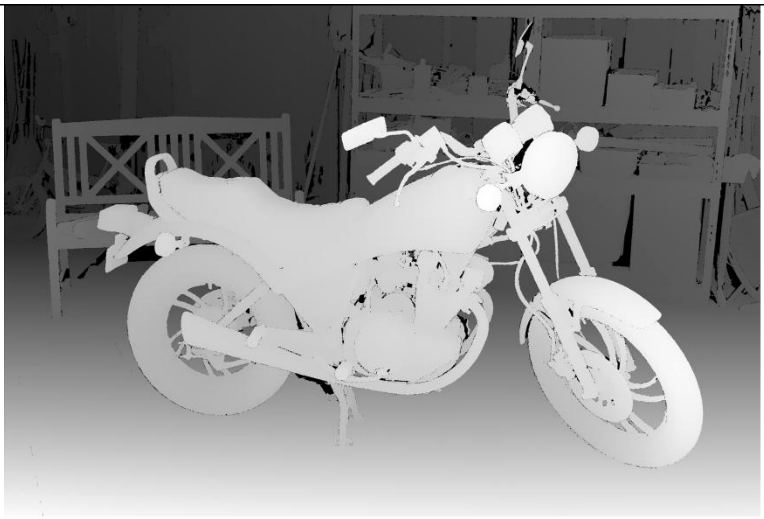
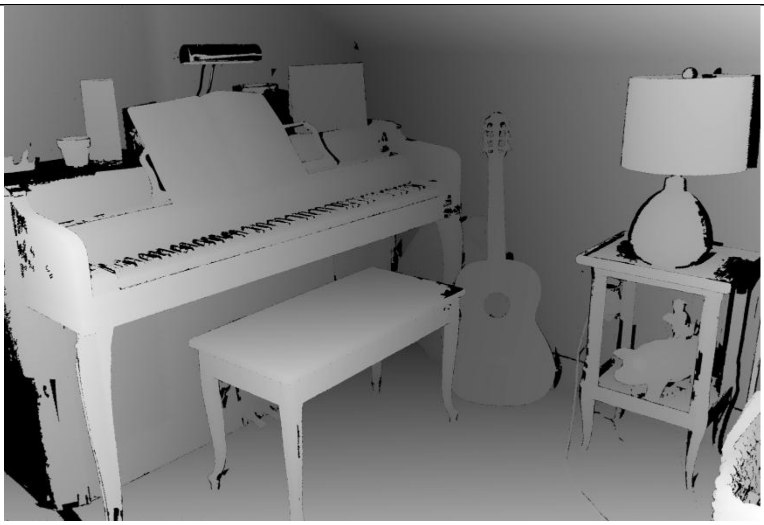
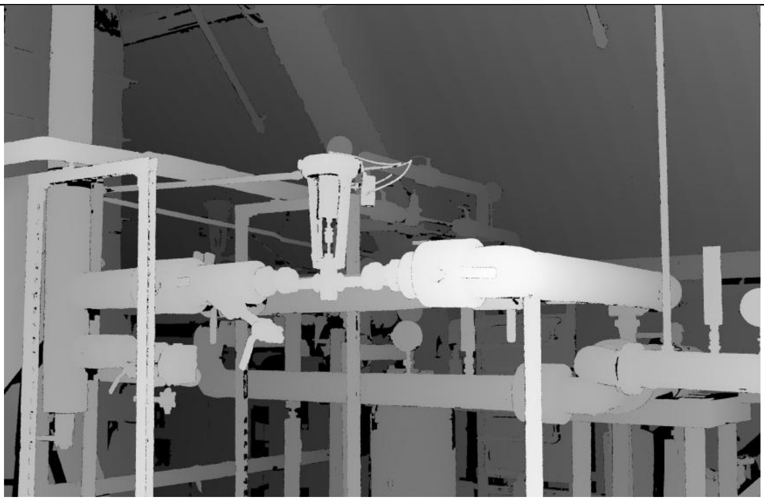
Figure 4.2.2.1. Two synthesized right images *Jadeplant* using BM.

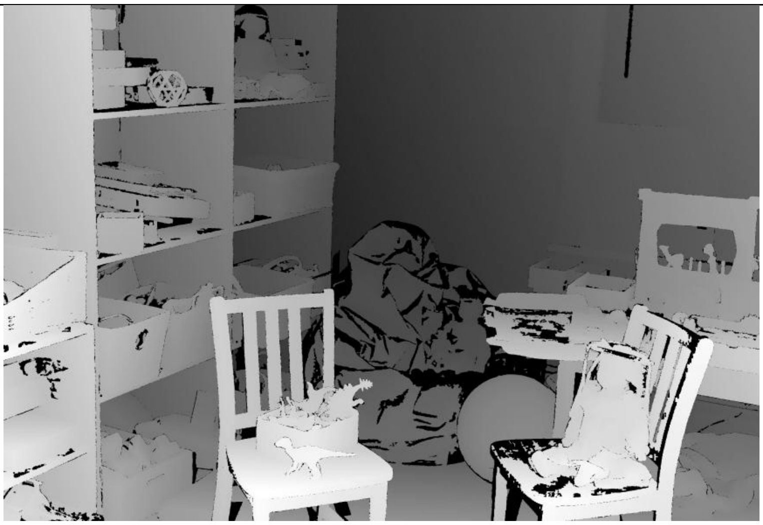
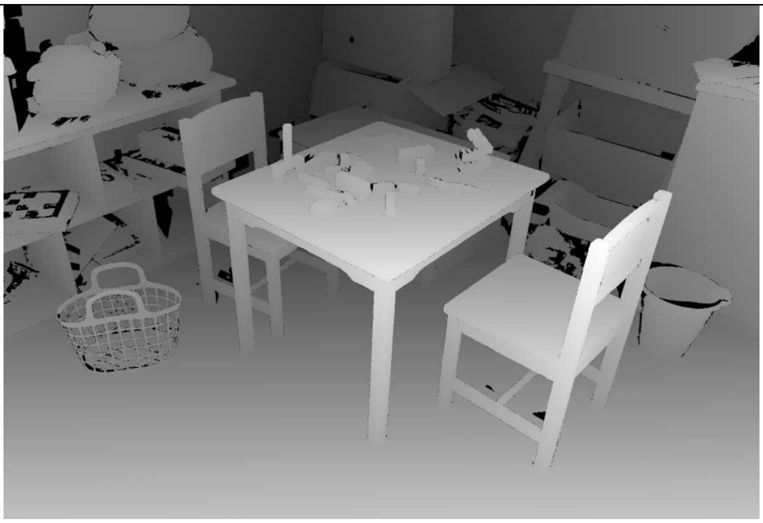
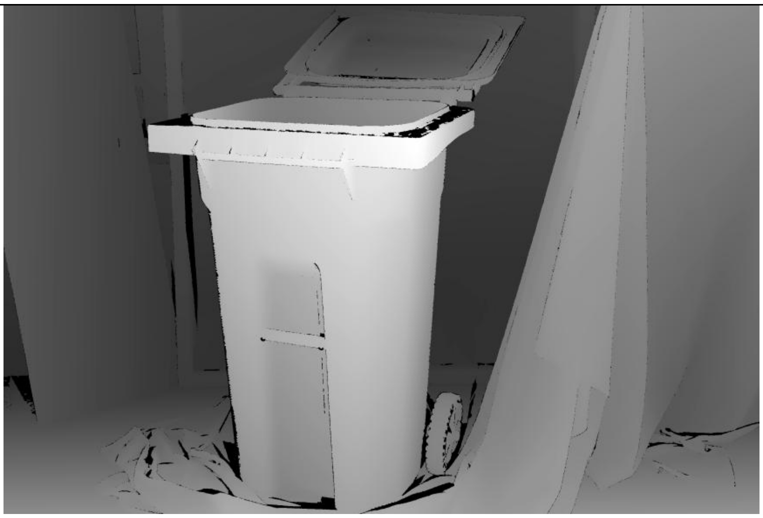
This case shows two visually different situations for synthesized images using different quality disparity maps, with an equal computed MSSIM value. For this particular case, the MSE is useful as a tie-breaker, where the proposed approach selects the image with lower MSE (*SADwindow=15*) among the two choices. This situation is tackled in (Dosselmann & Yang, 2011), where a relation between SSIM and MSE is shown empirically and formally. These similarities include one of the drawbacks of MSE in which for different perceived distortions a similar image quality value is computed.

4.2.3 Results comparison using *ground truth*

The stereo correspondence algorithm candidates selected for each stereo image pair in the past section presumably have the lowest error rate among the choices. This section presents a comparison with a linear correlation analysis between the MSSIM scores obtained using the proposed method and a *ground truth* error measuring using the BMP metric. Using the *ground truth* information available at Middlebury's website for each of the stereo image pairs in the dataset and each disparity map in section 4.1.1 the BMP metric was computed. Table 4.2.3.1 shows the ground truth data available at Middlebury's website (Scharstein et al., 2014). Table 4.2.3.2 shows the obtained MSSIM results in section 4.2.2 for each stereo image pair. Table 4.2.3.3 shows the computed BMP results for the dataset available at Middlebury's site.

Image	Ground truth disparity map
Adirondack	
Jadeplant	

Motorcycle	
Piano	
Pipes	

Playroom	
Playtable	
Recycle	

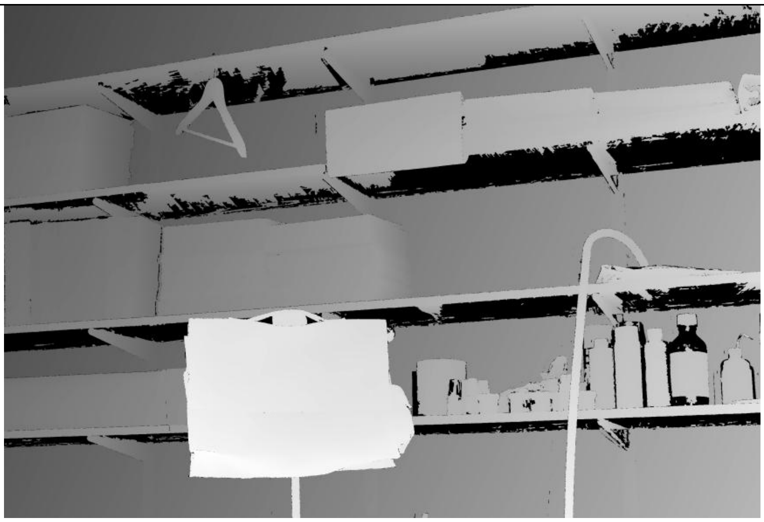
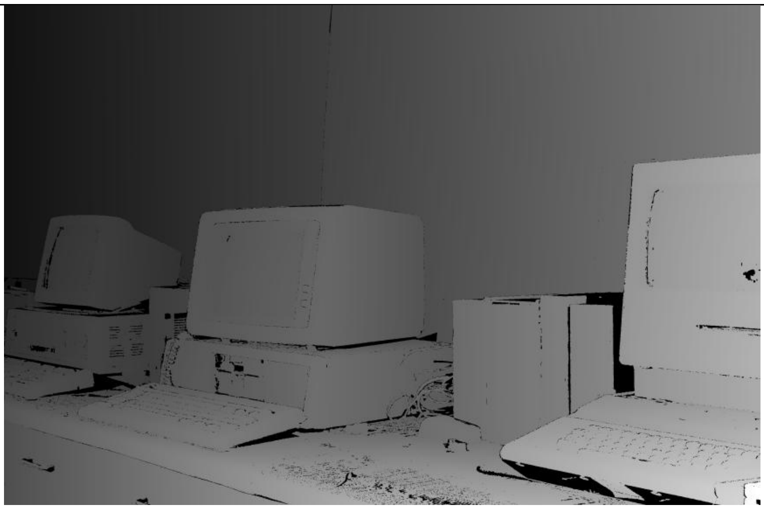
Shelves	
Vintage	

Table 4.2.3.1. *Ground truth* disparity maps for Middlebury dataset.

Algorithm	Adirondack	Jadeplant	Motorcycle	Piano	Pipes	Playroom	Playtable	Recycle	Shelves	Vintage
BM [5]	0.05	0.09	0.14	0.16	0.17	0.13	0.09	0.02	0.06	0.01
BM [11]	0.11	0.13	0.26	0.26	0.37	0.17	0.13	0.04	0.10	0.03
BM [15]	0.14	0.14	0.31	0.29	0.41	0.19	0.14	0.05	0.14	0.06
BM [29]	0.20	0.14	0.35	0.30	0.45	0.22	0.13	0.07	0.21	0.11
BM [45]	0.21	0.11	0.34	0.18	0.44	0.22	0.10	0.07	0.23	0.06
SGBM [7]	0.06	0.13	0.28	0.27	0.38	0.15	0.15	0.05	0.10	0.02
SGBM [9]	0.16	0.23	0.43	0.40	0.51	0.25	0.28	0.12	0.20	0.10
SGBM [11]	0.23	0.28	**0.52**	0.46	**0.57**	0.32	**0.35**	0.18	0.28	0.17
SGBM [15]	0.29	**0.31**	0.45	**0.50**	0.56	**0.36**	0.34	**0.25**	**0.33**	**0.21**
SGBM [17]	**0.30**	0.24	0.21	0.31	0.35	0.23	0.18	0.21	0.31	0.14

Table 4.2.3.2. MSSIM scores obtained with the proposed method according to section 4.2.2.

Algorithm	Adirondack	Jadeplant	Motorcycle	Piano	Pipes	Playroom	Playtable	Recycle	Shelves	Vintage
BM [5]	0.83	0.83	0.74	0.71	0.66	0.80	0.76	0.89	0.77	0.89
BM [11]	0.70	0.73	0.51	0.58	0.46	0.69	0.62	0.79	0.66	0.80
BM [15]	0.66	0.72	0.46	0.56	0.43	0.66	0.61	0.75	0.63	0.77
BM [29]	0.63	0.75	0.52	0.56	0.42	0.65	0.76	0.75	0.61	0.74
BM [45]	0.66	0.81	0.60	0.71	0.45	0.67	0.83	0.80	0.61	0.80
SGBM [7]	0.78	0.70	0.50	0.55	0.44	0.75	0.58	0.72	0.67	0.80
SGBM [9]	0.65	0.63	0.34	0.45	0.34	0.64	0.45	0.56	0.57	0.71
SGBM [11]	0.59	0.61	**0.30**	0.40	**0.32**	0.59	**0.43**	0.48	0.54	0.66
SGBM [15]	**0.54**	**0.60**	0.42	**0.39**	0.34	**0.56**	0.50	**0.44**	**0.52**	**0.64**
SGBM [17]	0.59	0.68	0.75	0.63	0.58	0.70	0.77	0.55	0.56	0.76

Table 4.2.3.3. Percentage of bad matched pixels obtained using the dataset *ground truth* data.

The data in tables 4.2.3.2 and 4.2.3.3 show that both methods agreed on the disparity maps selection with the presumable lower error rate in nine out of ten different scenes. Additionally, figure 4.2.3.1 shows a scatterplot for all of the data in tables 4.2.3.2 and 4.3.3.3 in order to compare the results obtained using the proposed approach and the BMP percentage commonly used in stereo correspondence evaluation. The data shows a strong negative correlation between the MSSIM results obtained with the proposed method and the BMP metric, with a Pearson's correlation coefficient $r = -0.902$.

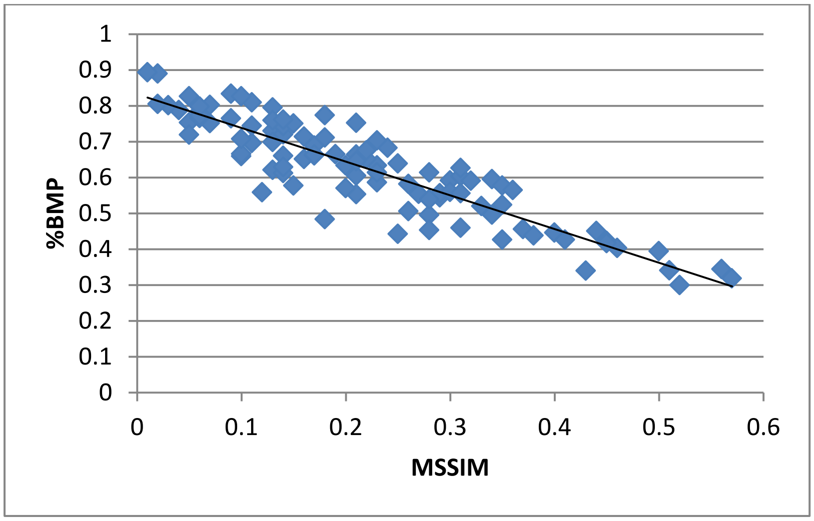


Fig 4.2.3.1. MSSIM vs BMP scatterplot.

The scatterplot in figure 4.2.3.1 allows concluding than in the presented dataset, the proposed method performs similarly to the widely used percentage of bad matched pixels which uses the ground truth information.

5. Final remarks

This document presents the results of a research oriented to the automated selection of stereo correspondence algorithms in absence of ground truth. The approach presented is based on the prediction error disparity maps evaluation method that uses only two images to assess the quality and perform a selection of a stereo correspondence algorithm. The concepts required to understand a stereo vision system and the methods for assessing the quality of estimated disparity maps are described.

Two systematic reviews are developed in order to characterize the state-of-art techniques for assessing the quality of disparity maps and to perform the 3D image warping required in prediction error approaches. Firstly, a taxonomy of disparity maps quality assessment methods is defined and a conceptual comparison of the prediction error methods is accomplished. Secondly, a description of the different 3D image warping techniques is presented along with the respective considerations of using them in the context of this thesis.

The method proposed in this thesis is performed by using three main processes: disparity estimation, 3D image warping, and 2D image quality assessment. First, the disparity maps are estimate using a fixed set of stereo correspondence algorithms and parameters. Then, a synthesized right image is warped using the input left image and each assessed disparity map. Finally, the similarity between the synthesized images and the reference right image is measured using the structural similarity index (SSIM). The scores obtained from the process are used to select the disparity map with presumably the lower error rate.

Two tests were conducted to prove the functionality of the proposed method. The Middlebury dataset in its third vision along with two available stereo correspondence algorithms implementations on the opencv library were processed using the proposed selection method. Results show that the developed prototype for the proposed method is functional and shows a strong correlation with a *ground truth* based disparity map evaluation approach for the processed data.

As future work, the proposed method in this thesis can be embed as the core of a robust stereo vision system. When stereo vision is applied to real world applications small changes on the assessed scene contents are a common issue, since is well known that the quality of a disparity map estimated by a stereo correspondence algorithm depends of the contents of the scene. Using the approach proposed in this thesis, a stereo system can be developed to compute several different stereo correspondence algorithms with different parameters in near real-time, and select the best prospect among the available disparity maps. This could contribute to ease the implementation of stereo vision systems into fields such as industry by adding adaptability and robustness properties.

6. Publications

The paper “Stereo Correspondence Evaluation Methods: A Systematic Review” has been published for the 11th International Symposium on Visual Computing (ISVC’15) by Springer-Verlag in the Lecture Notes in Computer Science (LNCS) series as a result of the research presented in this thesis.

Currently, efforts are being made in order to publish a detailed implementation of the proposed method on this thesis as a future academic product of this work.

7. References

Aguilar, M. A., del Mar Saldana, M., & Aguilar, F. J. (2014). Generation and Quality Assessment of Stereo-Extracted DSM From GeoEye-1 and WorldView-2 Imagery. In *IEEE Transactions on Geoscience and Remote Sensing* (Vol. 52, pp. 1259–1271)

Bhola, V. K., Sharma, T., & Bhatnagar, J. (n.d.). Image Quality Assessment Techniques.

Birchfield, S., & Tomasi, C. (1998). A pixel dissimilarity measure that is insensitive to image sampling. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, *20*(4), 401–406.

Bradski, G. R., & Kaehler, A. (2008). *Learning OpenCV: computer vision with the OpenCV library*.

Cabezas, I. (2013). Evaluation of disparity maps, Doctoral thesis. Universidad del Valle.

Cabezas, I., Padilla, V., & Trujillo, M. (2011). A measure for accuracy disparity maps evaluation. In *Progress in Pattern Recognition, Image Analysis, Computer Vision, and Applications* (pp. 223–231). Springer.

Cabezas, I., Padilla, V., & Trujillo, M. (2012). BMPRE: An Error measure for evaluating disparity maps (Vol. 2, pp. 1051–1055). Presented at the International Conference on Signal Processing Proceedings, ICSP.

Choi, H., Seo, Y., Yoo, J., & Kim, D. (2013). Multi-View Stereoscopic Image Synthesis Algorithm for 3DTV. In H.-K. Jung, J. T. Kim, T. Sahama, & C.-H. Yang (Eds.), *Future Information Communication Technology and Applications* (Vol. 235, pp. 509–517). Dordrecht: Springer Netherlands.

Devernay, F., & Duchêne, S. (2010). New view synthesis for stereo cinema by hybrid disparity remapping. In *Image Processing (ICIP), 2010 17th IEEE International Conference on* (pp. 5–8). IEEE.

Devernay, F., & Peon, A. R. (2010). Novel view synthesis for stereoscopic cinema: detecting and removing artifacts. In *Proceedings of the 1st international workshop on 3D video processing* (pp. 25–30). ACM.

Dosselmann, R., & Yang, X. D. (2011). A comprehensive assessment of the structural similarity index. *Signal, Image and Video Processing*, *5*(1), 81–91.

Fehn, C. (2004). Depth-image-based rendering (DIBR), compression, and transmission for a new approach on 3D-TV. In *Electronic Imaging 2004* (pp. 93–104). International Society for Optics and Photonics.

Fuhr, G., Fickel, G. P., Dal'Aqua, L. P., Jung, C. R., Malzbender, T., & Samadani, R. (2013). An evaluation of stereo matching methods for view interpolation. In *Image Processing (ICIP), 20th IEEE International Conference on* (pp. 403–407). IEEE.

Geetha Ramachandran, & Markus Rupp (Eds.). (2012). Multiview synthesis from stereo views.

Geiger, A., Lenz, P., & Urtasun, R. (2012). Are we ready for autonomous driving? the kitti vision benchmark suite. In *Computer Vision and Pattern Recognition (CVPR), IEEE Conference on* (pp. 3354–3361). IEEE.

Gong, M., Yang, R., Wang, L., & Gong, M. (2007). A performance study on different cost aggregation approaches used in real-time stereo matching. *International Journal of Computer Vision*, *75*(2), 283–296.

Haeusler, R., & Klette, R. (2010). Benchmarking Stereo Data (Not the Matching Algorithms). In *DAGM-Symposium* (pp. 383–392). Springer.

Haeusler, R., & Klette, R. (2012). Evaluation of stereo confidence measures on synthetic and recorded image data. In *Informatics, Electronics & Vision (ICIEV), International Conference on* (pp. 963–968). IEEE.

Haeusler, R., Nair, R., & Kondermann, D. (2013). Ensemble learning for confidence measures in stereo vision. In *Computer Vision and Pattern Recognition (CVPR), IEEE Conference on* (pp. 305–312). IEEE.

Hamilton, O. K., Breckon, T. P., Bai, X., & Kamata, S. (2013). A foreground object based quantitative assessment of dense stereo approaches for use in automotive environments. In *Image Processing (ICIP), 20th IEEE International Conference on* (pp. 418–422). IEEE.

Hermann, S., Morales, S., & Klette, R. (2011). Half-resolution semi-global stereo matching. In *Intelligent Vehicles Symposium (IV), IEEE* (pp. 201–206). IEEE.

Hirschmuller, H. (2008). Stereo Processing by Semiglobal Matching and Mutual Information. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *30*(2), 328–341.

Hirschmuller, H., & Scharstein, D. (2007). Evaluation of cost functions for stereo matching. In *Computer Vision and Pattern Recognition, 2007. CVPR’07. IEEE Conference on* (pp. 1–8). IEEE.

Hirschmuller, H., & Scharstein, D. (2009). Evaluation of Stereo Matching Costs on Images with Radiometric Differences. In *IEEE Transactions on Pattern Analysis and Machine Intelligence* (Vol. 31, pp. 1582–1599).

Hsiao, S.-F., Cheng, J.-W., Wang, W.-L., & Yeh, G.-F. (2012). Low latency design of depth-image-based rendering using hybrid warping and hole-filling. In *Circuits and Systems (ISCAS), 2012 IEEE International Symposium on* (pp. 608–611). IEEE.

Huang, Y.-H., Huang, T.-K., Huang, Y.-H., Chen, W.-C., & Chuang, Y.-Y. (2012). Warping-Based Novel View Synthesis from a Binocular Image for Autostereoscopic Displays (pp. 302–307). IEEE.

Jung, C., Jiao, L., Oh, Y., & Kim, J. K. (2010). Depth-preserving DIBR based on disparity map over T-DMB. *Electronics Letters*, *46*(9), 628–629.

Keller, C. G., Enzweiler, M., & Gavrila, D. M. (2011). A new benchmark for stereo-based pedestrian detection. In *Intelligent Vehicles Symposium (IV)* (pp. 691–696). IEEE.

Kelly, P. (2007). Pedestrian detection and tracking using stereo vision techniques. Dublin City University.

Kelly, P., O’Connor, N. E., & Smeaton, A. F. (2008). A Framework for Evaluating Stereo-Based Pedestrian Detection Techniques. In *IEEE Transactions on Circuits and Systems for Video Technology* (Vol. 18, pp. 1163–1167).

Kim, T. (2010). Intermediate view synthesis algorithm using mesh clustering for rectangular multiview camera system. *Optical Engineering*, *49*(2), 027002.

Kitchenham, B. (2004). Procedures for performing systematic reviews. In *Keele, UK, Keele University* (Vol. 33, pp. 1–26).

Kogler, J., Eibensteiner, F., Humenberger, M., Gelautz, M., & Scharinger, J. (2013). Ground Truth Evaluation for Event-Based Silicon Retina Stereo Data (pp. 649–656). IEEE.

Kondermann, D., Nair, R., Meister, S., Mischler, W., Güssefeld, B., Honauer, K., … Jähne, B. (2015). Stereo Ground Truth with Error Bars. In *Computer Vision–ACCV 2014* (pp. 595–610). Springer.

Kostlivá, J., Čech, J., & others. (2007). Feasibility boundary in dense and semi-dense stereo matching. In *Computer Vision and Pattern Recognition. CVPR'07. IEEE Conference on* (pp. 1–8). IEEE.

Leclercq, P., & Morris, J. (2003). Robustness to noise of stereo matching. In *Image Analysis and Processing, 2003. Proceedings. 12th International Conference on* (pp. 606–611). IEEE.

Lee, G.-C., & Yoo, J. (2015). Disparity refinement near the object boundaries for virtual-view quality enhancement. *Journal of Electrical Engineering and Technology*, *10*(5), 2189–2196.

Leibe, B., Cornelis, N., Cornelis, K., & Van Gool, L. (2007). Dynamic 3d scene analysis from a moving vehicle. In *Computer Vision and Pattern Recognition. CVPR'07. IEEE Conference on* (pp. 1–8). IEEE.

Lei, L., Chen, Z., & Shi, J. (2014). A hole-filling algorithm based on pixel labeling for DIBR (Vol. 9284). Presented at the Proceedings of SPIE - The International Society for Optical Engineering.

Liu, W., Zhang, D., Cui, M., & Ding, J. (2015). *An enhanced depth map based rendering method with directional depth filter and image inpainting*.

Lü, C., Wang, H., Ren, H., & Shen, Y. (2010). Virtual View Synthesis for Multi-view 3D Display (pp. 444–446). IEEE.

Malpica, W., & Bovik, A. C. (2008). Range Image Quality Assessment by Structural Similarity. In *Encyclopedia of Multimedia* (pp. 757–762).

Manap, N. A., & Soraghan, J. J. (2011). Novel view synthesis based on depth map layers representation. In *3DTV Conference: The True Vision-Capture, Transmission and Display of 3D Video (3DTV-CON), 2011* (pp. 1–4). IEEE.

Manap, N. A., & Soraghan, J. J. (2014). Disparity depth map layers representation for image view synthesis. *Journal of Telecommunication, Electronic and Computer Engineering*, *6*(1), 1–8.

Milani, S., Ferrario, D., & Tubaro, S. (2013). No-reference quality metric for depth maps. In *Image Processing (ICIP), 20th IEEE International Conference on* (pp. 408–412). IEEE.

Morales, S., & Klette, R. (2009). A third eye for performance evaluation in stereo sequence analysis. In *Computer Analysis of Images and Patterns* (pp. 1078–1086). Springer.

Morales, S., & Klette, R. (2011). Ground truth evaluation of stereo algorithms for real world applications. In *Computer Vision–ACCV 2010 Workshops* (pp. 152–162). Springer.

Morales, S., Vaudrey, T., & Klette, R. (2009). Robustness evaluation of stereo algorithms on long stereo sequences. In *Intelligent Vehicles Symposium, IEEE* (pp. 347–352). IEEE.

Neilson, D., & Yang, Y.-H. (2008). Evaluation of constructable match cost measures for stereo correspondence using cluster ranking. In *Computer Vision and Pattern Recognition, CVPR. IEEE Conference on* (pp. 1–8). IEEE.

Nielsen, M., Andersen, H. J., Slaughter, D. C., & Granum, E. (2007). Ground truth evaluation of computer vision based 3D reconstruction of synthesized and real plant images. In *Precision Agriculture* (Vol. 8, pp. 49–62).

Petersen, K., Feldt, R., Mujtaba, S., & Mattsson, M. (2008). Systematic Mapping Studies in Software Engineering (pp. 68–77). Presented at the Proceedings of the 12th International Conference on Evaluation and Assessment in Software Engineering, Italy: British Computer Society.

Rhee, S.-M., Choi, J., & Choi, S. (2010). Accurate stereo view synthesis for an autostereoscopic 3D display.

Riechert, C., Zilly, F., Müller, M., & Kauff, P. (2012). Advanced interpolation filters for depth image based rendering. In *3DTV-Conference: The True Vision-Capture, Transmission and Display of 3D Video (3DTV-CON), 2012* (pp. 1–4). IEEE.

Scharstein, D. (1999). *View synthesis using stereo vision*. Berlin ; New York: Springer.

Scharstein, D., Hirschmüller, H., Kitajima, Y., Krathwohl, G., Nešić, N., Wang, X., & Westling, P. (2014). High-resolution stereo datasets with subpixel-accurate ground truth. In *Pattern Recognition* (pp. 31–42). Springer.

Scharstein, D., & Szeliski, R. (2002). A taxonomy and evaluation of dense two-frame stereo correspondence algorithms. In *International journal of computer vision* (Vol. 47, pp. 7–42).

Scharstein, D., & Szeliski, R. (2003). High-accuracy stereo depth maps using structured light. In *Computer Vision and Pattern Recognition, 2003. Proceedings. 2003 IEEE Computer Society Conference on* (Vol. 1, pp. I–195). IEEE.

Sellent, A., & Wingbermühle, J. (2012). Quality assessment of non-dense image correspondences. In *Computer Vision–ECCV 2012. Workshops and Demonstrations* (pp. 114–123). Springer.

Shin, B.-S., Caudillo, D., & Klette, R. (2015). Evaluation of two stereo matchers on long real-world video sequences. In *Pattern Recognition* (Vol. 48, pp. 1113–1124).

Shin, I.-Y., & Ho, Y.-S. (2012). Virtual viewpoint disparity estimation and convergence check for real-time view synthesis. In *Advances in Image and Video Technology* (pp. 121–131). Springer.

Steingrube, P., Gehrig, S. K., & Franke, U. (2009). Performance evaluation of stereo algorithms for automotive applications. In *Computer vision systems* (pp. 285–294). Springer.

Szeliski, R. (1999). Prediction error as a quality metric for motion and stereo. In *Computer Vision. The Proceedings of the Seventh IEEE International Conference on* (Vol. 2, pp. 781–788). IEEE.

Szeliski, R. (2010). *Computer vision: algorithms and applications*. Springer Science & Business Media.

Szeliski, R., & Zabih, R. (2000). An experimental comparison of stereo algorithms. In *Vision algorithms: theory and practice* (pp. 1–19). Springer.

Tombari, F., Mattoccia, S., & Di Stefano, L. (2010). Stereo for robots: quantitative evaluation of efficient and low-memory dense stereo algorithms. In *Control Automation Robotics & Vision (ICARCV), 11th International Conference on* (pp. 1231–1238). IEEE.

Tran, L. C., Pal, C. J., & Nguyen, T. Q. (2010). View synthesis based on conditional random fields and graph cuts. In *Image Processing (ICIP), 2010 17th IEEE International Conference on* (pp. 433–436). IEEE.

Tran, L., Khoshabeh, R., Jain, A., Pal, C., & Nguyen, T. (2011). Spatially consistent view synthesis with coordinate alignment. In *Acoustics, Speech and Signal Processing (ICASSP), 2011 IEEE International Conference on* (pp. 905–908). IEEE.

vanderMark, W., & Gavrila, D. M. (2006). Real-Time Dense Stereo for Intelligent Vehicles. In *IEEE Transactions on Intelligent Transportation Systems* (Vol. 7, pp. 38–50).

Vandewalle, P., & Varekamp, C. (2014). Disparity map quality for image-based rendering based on multiple metrics. In *3D Imaging (IC3D), International Conference on* (pp. 1–5). IEEE.

Varekamp, C., Hinnen, K., & Simons, W. (2013). Detection And Correction Of Disparity Estimation Errors Via Supervised Learning. In *3D Imaging (IC3D), 2013 International Conference on* (pp. 1–7). IEEE.

Vargas, C., Cabezas, I., & Branch, J. W. (2015). Stereo Correspondence Evaluation Methods: A Systematic Review. In G. Bebis, R. Boyle, B. Parvin, D. Koracin, I. Pavlidis, R. Feris, … G. Weber (Eds.), *Advances in Visual Computing* (Vol. 9475, pp. 102–111). Cham: Springer International Publishing.

Wang, Z., Bovik, A. C., Sheikh, H. R., & Simoncelli, E. P. (2004). Image Quality Assessment: From Error Visibility to Structural Similarity. *IEEE Transactions on Image Processing*, *13*(4), 600–612.

Wang, Z.-F., & Zheng, Z.-G. (2008). A region based stereo matching algorithm using cooperative optimization. In *Computer Vision and Pattern Recognition, 2008. CVPR 2008. IEEE Conference on* (pp. 1–8). IEEE.

Wedel, A., Rabe, C., Vaudrey, T., Brox, T., Franke, U., & Cremers, D. (2008). *Efficient dense scene flow from sparse or dense stereo data*. Springer.

Woodward, A., Leclercq, P., Delmas, P., & Gimel’farb, G. (2006). Generation of an Accurate Facial Ground Truth for Stereo Algorithm Evaluation. In *Computer Vision and Graphics* (pp. 534–539). Springer.

Xiaoyan Hu, & Mordohai, P. (2012). A Quantitative Evaluation of Confidence Measures for Stereo Vision. In *IEEE Transactions on Pattern Analysis and Machine Intelligence* (Vol. 34, pp. 2121–2133).

Yang, L., Yendo, T., Panahpour, M., Fujii, T., & Tanimoto, M. (2010). View synthesis using probabilistic reliability reasoning for FTV.

Yang, L., Yendo, T., Tehrani, M. P., Fujii, T., & Tanimoto, M. (2010a). Error supression in view synthesis using reliability reasoning for FTV. In *3DTV-Conference: The True Vision-Capture, Transmission and Display of 3D Video (3DTV-CON), 2010* (pp. 1–4). IEEE.

Yang, L., Yendo, T., Tehrani, M. P., Fujii, T., & Tanimoto, M. (2010b). Probabilistic reliability based view synthesis for FTV. In *Image Processing (ICIP), 2010 17th IEEE International Conference on* (pp. 1785–1788). IEEE.

Yao, S.-J., Wang, L.-H., Lin, C.-L., & Zhang, M. (2015). *Real-time stereo to multi-view conversion system based on adaptive meshing*.

Yinghua Shen, Chaohui Lu, Pin Xu, & Lili Xu. (2011). Objective Quality Assessment of Noised Stereoscopic Images (pp. 745–747). IEEE.

Zhang, Z., Hou, C., Shen, L., & Yang, J. (2009). An Objective Evaluation for Disparity Map Based on the Disparity Gradient and Disparity Acceleration (pp. 452–455).

Zhu, S., Li, Z., & Yu, Y. (2014). Virtual view synthesis using stereo vision based on the sum of absolute difference. *Computers and Electrical Engineering*, *40*(8), 236–246.