12-2008

MRF Stereo Matching with Statistical Estimation of Parameters

Mohammad Shafikul Huq
University of Tennessee - Knoxville

To the Graduate Council:

I am submitting herewith a dissertation written by Mohammad Shafikul Huq entitled "MRF Stereo Matching with Statistical Estimation of Parameters." I have examined the final electronic copy of this dissertation for form and content and recommend that it be accepted in partial fulfillment of the requirements for the degree of Doctor of Philosophy, with a major in Electrical Engineering.

Mongi A. Abidi, Major Professor

We have read this dissertation and recommend its acceptance:

Paul B. Crilly, Seddik M. Djouadi, Frank M. Guess, Andreas Koschan

Accepted for the Council:
Carolyn R. Hodges

Vice Provost and Dean of the Graduate School

(Original signatures are on file with official student records.)

To the Graduate Council:

I am submitting herewith a dissertation written by Mohammad Shafikul Huq entitled "MRF Stereo Matching with Statistical Estimation of Parameters." I have examined the final electronic copy of this dissertation for form and content and recommend that it be accepted in partial fulfillment of the requirements for the degree of Doctor of Philosophy, with a major in Electrical Engineering.

Mongi A. Abidi , Major Professor

We have read this dissertation
and recommend its acceptance:

Paul B. Crilly

Seddik M. Djouadi

Frank M. Guess

Andreas Koschan

Accepted for the Council:

Carolyn R. Hodges
Vice Provost and Dean of the Graduate School

(Original signatures are on file with official student records.)

MRF Stereo Matching with Statistical Estimation of Parameters

A Dissertation Proposal
Presented for the
Doctor of Philosophy
Degree
The University of Tennessee at Knoxville

Mohammad Shafikul Huq
Nov 12, 2008

Abstract

For about the last ten years, stereo matching in computer vision has been treated as a combinatorial optimization problem. Assuming that the points in stereo images form a Markov Random Field (MRF), a variety of combinatorial optimization algorithms has been developed to optimize their underlying cost functions. In many of these algorithms, the MRF parameters of the cost functions have often been manually tuned or heuristically determined for achieving good performance results. Recently, several algorithms for statistical, hence, automatic estimation of the parameters have been published. Overall, these algorithms perform well in labeling, but they lack in performance for handling discontinuity in labeling along the surface borders.

In this dissertation, we develop an algorithm for optimization of the cost function with automatic estimation of the MRF parameters – the data and smoothness parameters. Both the parameters are estimated statistically and applied in the cost function with support of adaptive neighborhood defined based on color similarity. With the proposed algorithm, discontinuity handling with higher consistency than of the existing algorithms is achieved along surface borders. The data parameters are pre-estimated from one of the stereo images by applying a hypothesis, called noise equivalence hypothesis, to eliminate interdependency between the estimations of the data and smoothness parameters. The smoothness parameters are estimated applying a combination of maximum likelihood and disparity gradient constraint, to eliminate nested inference for the estimation. The parameters for handling discontinuities in data and smoothness are defined statistically as well. We model cost functions to match the images symmetrically for improved matching performance and also to detect occlusions. Finally, we fill the occlusions in the disparity map by applying several existing and proposed algorithms and show that our best proposed segmentation based least squares algorithm performs better than the existing algorithms.

We conduct experiments with the proposed algorithm on publicly available ground truth test datasets provided by the Middlebury College. Experiments show that results better than the existing algorithms' are delivered by the proposed algorithm having the MRF parameters estimated automatically. In addition, applying the parameter estimation technique in existing stereo matching algorithm, we observe significant improvement in computational time.

Table of contents

List of figures

List of tables

List of publications

Book Chapter:

S. Huq, B. Abidi, S. Kong, and M. Abidi. Chapter 2: A Survey on 3D Modeling of Human Faces for Face Recognition. *3D Imaging for Safety and Security*, pp. 25–67. Publisher: *Springer*, Jul 2007; ISBN: 978-1-4020-6181-3

Conferences:

(1). *Shafik Huq*, Andrea Koschan, Besma Abidi, and Mongi Abidi, "Efficient BP Stereo with Automatic Parameter Estimation," to appear in the *Proc. of IEEE 15th Int'l Conf. on Image Processing* (ICIP), Oct 2008.
(2). *Shafik Huq*, Andreas Koschan, Besma Abidi, and Mongi Abidi, "MRF Stereo with Statistical Estimation of Parameters," *Proc. of IEEE 4th Int'l Symposium on 3D Data Processing, Visualization, and Transmission* (3DPVT), Atlanta, Jun 2008.
(3). *Shafik Huq*, Besma Abidi, David Page, and Mongi Abidi, J. Frafjord, and S. Deckanich, "Inspection of Fracture Surfaces using 3D from Stereo Images of Large Chamber SEM," *Int'l Conf. on Microscopy and Microanalysis* (ICMM), Florida, Aug 5–9, 2007.
(4). *Shafik Huq*, Besma Abidi, and Mongi Abidi, "Stereo-based 3D Face Modeling using Annealing in Local Energy Minimization," *IEEE 14th Int'l Conf. on Image Analysis and Processing* (ICIAP), Modena, Italy, Sep 10–13, 2007.
(5). *S. Huq*, B. Abidi, C. Kammerud, M. Abidi, J. Frafjord, and S. Deckanich, "3D Measurements of Wear on Machining Tools Using a Confocal Microscope," *Int'l Conf. on Microscopy and Microanalysis jointly with Int'l Metallographic Society* (ICIMS), Vol. 34, No. 2, Aug 2006.
(6). C. Kammerud, B. Abidi, *S. Huq*, and M. Abidi, "3D Nanovision for the Inspection of Micro-Electro-Mechanical Systems," *IEEE Int'l Conf. on Electronics, Circuits, and Systems* (ICECS), Gammarth, Tunisia, Dec 2005.
(7). B. Abidi, *S. Huq*, and M. Abidi, "Fusion of Visual, Thermal, and Range as a Solution to Illumination and Pose Restrictions in Face Recognition," *Proc. of IEEE Carnahan Conf. on Security Technology* (CCST), Albuquerque, NM, pp. 325–330, Oct 2004.
(8). *Shafik Huq*, Besma Abidi, Ardeshir Goshtasby, and Mongi Abidi, "Stereo Matching with Energy Minimizing Snake Grid for 3D Face Modeling," *Proc. of SPIE Defense and Security Symposium* (SDSS), Vol. 5405, pp. 530–536, Apr 2004.

Topical Meetings:

(1). *Shafik Huq*, Besma Abidi, David Page, Mongi Abidi, J. Frafjord, and S. Deckanich, "3D Modeling from Large Chamber SEM Stereo Images for Micro-scale Surface Inspection and Characterization," *2nd Topical Int'l Meeting on Emergency Preparedness & Response and Robotic & Remote Systems*, NM, USA, Mar 2008.
(2). W. Hao, *S. Huq*, D. Page, B. Abidi, A. Koschan, and M. Abidi, "Nano-Scale 3D Metrology for Surface Characterization and Inspection of High-Precision Manufactured Components," *ANS/ENS International Meeting*, Washington, DC, Nov 11–15, 2007.
(3) *S. Huq*, B. Abidi, D. Page, M. Abidi, J. Frafjord, and S. Deckanich, "Nano-scale Imaging Research," *Poster in National Nuclear Security Administration* (NNSA), Sep 2006.

1 Introduction

In the process of visual perception of depth, human eyes form a triangle with the scene. This process of visualization of the 3D world is called triangulation. The key idea in triangulation is that, if an object is viewed with only one eye open at a time, the object is seen to be shifted. The amount of shifting is inversely proportional to the distance of the object from the location of the eyes. For about the last four decades, continuous effort has been provided by the computer vision scientists to implement triangulation and make a computer see like human. In the implementation, cameras (i.e. the eyes) are connected to a computer, images of a scene are captured, and then the scene is reconstructed into 3D from the images. This technique for 3D reconstruction is known as stereo vision, where one needs at least two images taken from two different viewpoints with respect to the scene. The number of cameras could be two or more with the images captured simultaneously at the same point of time. Alternatively, one camera can capture the images sequentially over time if the scene to be reconstructed remains still.

In Figure 1.1, a binocular stereo system is shown and the triangulation technique is illustrated. Figure 1.1 (a) shows the stereo system with two cameras placed some distance apart. Figure 1.2 shows a real stereo image pair. Note that the cameras could be oriented and arbitrarily placed, but they are focused on the same scene so that a significant overlap in the images is obtained. The portion of the scene which is visible in both the cameras can only be reconstructed into 3D. Figure 1.1 (b) illustrates how the depth of a scene point relates to the abovementioned shifting in the triangulation process. C_L and C_R are the camera centers. From geometry of the drawing, notice that the projections, p_f and p_c of the scene points P_f and P_c respectively, shift to the left in the right camera CCD plane (also known as the image plane) with respect to their location in the left camera CCD plane. This is true for any point in the left image that the point will shift to the left in the right image. Additionally, the condition, $x(p_{cL}) - x(p_{cR}) > x(p_{fL}) - x(p_{fR})$, holds, where $x(p)$ is the x-coordinate of a point, p; p_{cR} and p_{fR} are the projections of the scene points P_f and P_c on the right CCD plane. P_c is closer to the cameras than P_f. $x(p_{cL}) - x(p_{cR})$ is called the disparity of p_{cL}. Thus, in stereo vision a point located farther than other points has disparity smaller than the other points'.

Essentially, two problems need to be solved to reconstruct a scene using stereo technique. One is matching of the images, where for a point in one of the images disparity in the other image is determined. The other is 3D reconstruction of the scene points applying 2D image coordinates of the matched image points. The later problem is related to geometric calibration of the cameras that participate into acquisition of the

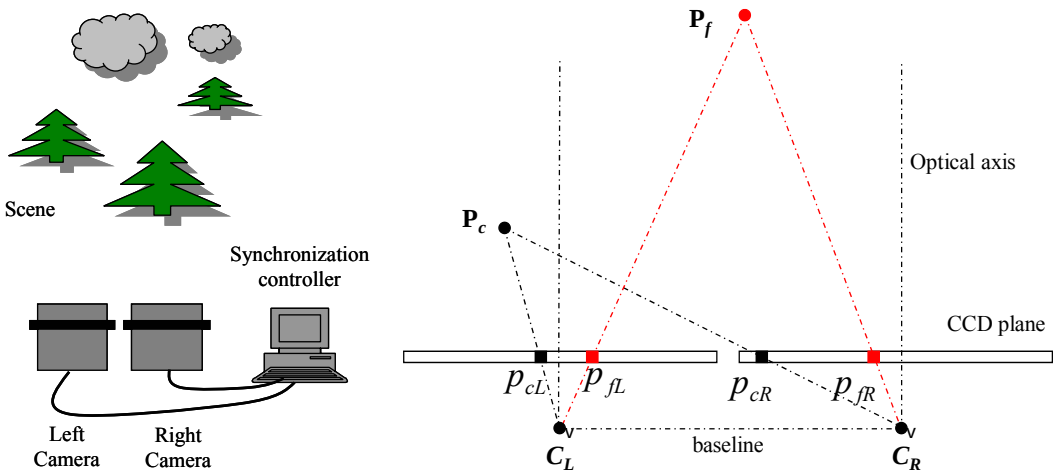


Figure 1.1: (a) A stereo setup; (b) demonstration of disparity depth relation with projections of the points on the CCD planes

stereo images. In this dissertation, we focus on solving the former problem, the stereo matching, which is proved to be very difficult. We consider the binocular stereo, where two cameras are located at two viewpoints on a line called baseline. We call the images, captured by the two cameras, the left and right stereo images. Then, more specifically, stereo matching can be described as a problem of matching the sites (pixels, features, or image regions) of the left image to their corresponding sites in the right image. Hundreds of algorithms have been published in the literature so far with new algorithms continuously developed and published. In June 2007, a search with the phrase "stereo matching" resulted about 514 hits only in the IEEE site.

Figure 1.2: A stereo pair with left and right images taken from two different viewpoints

1.1 Background

Any stereo matching problem follows several general approaches towards a solution. One of these approaches is window based, where, a window centering at a pixel to be matched is taken in the left image and matched (often with cross-correlation or sum of absolute differences) in the right image. There are two reasons behind considering a window instead of only the pixel itself. First, stereo matching is ambiguous. One pixel in the left image could match equally well with many pixels in the right image. A window is often able to reduce this ambiguity by enclosing any gradient existing in the neighborhood of the pixel. Second, individual pixels are more or less corrupted with noise. Matching cost aggregated on a window suppresses the effect of noise in the matching score; the price one has to pay is an increased computational cost and lost details in the reconstruction. The window based approaches assume that object surfaces are locally planar and also parallel to the image plane of the camera. Therefore, details in the reconstruction are lost, i.e. accuracy in matching is sacrificed, in image regions that have depth discontinuities, are unparallel with the image planes, or are taken with cameras placed large baseline length apart. In the last case, image regions are visibly distorted due to perspective foreshortening. Due to all these demerits, latest algorithms have focused on matching stereo images pixel by pixel to provide a dense and detailed reconstruction of the scene. However, in pixel level, ambiguity in matching is increased.

Both window and pixel based approaches apply a prior knowledge called spatial coherence to resolve the ambiguity. Spatial coherence says that neighboring pixels should have similar disparities, i.e., disparity map in a neighborhood should be smooth. Therefore, one considers two costs in matching a site (pixel): one is the data cost, related to data likelihood and the other is smoothness cost, related to spatial coherence or prior. Combining the data and smoothness costs together into a single cost to define a matching score has made stereo matching an optimization problem, which has evolved into a category of algorithms known as energy minimizing algorithms. The goal of these algorithms is to minimize the energy cost aggregated over all the pixels of the left image. Two major tasks are involved in achieving this goal: one, modeling of the cost function that accounts for the cost of a site (pixel or window) and two, optimization of the cost aggregated over all sites. In the modeling phase, one or more energy parameters, also known as the free parameters or temperatures, are taken into account to balance between the data and smoothness costs so that the resulting disparity map neither suffer from lack of smoothness nor become over-smoothed. Additionally, a few other energy parameters are defined and set to appropriate values so that discontinuities in the data and smoothness costs are handled properly. The energy minimizing algorithms are more specifically called MRF algorithms when the related cost functions are modeled in probabilistic framework.

1.2 State of the art

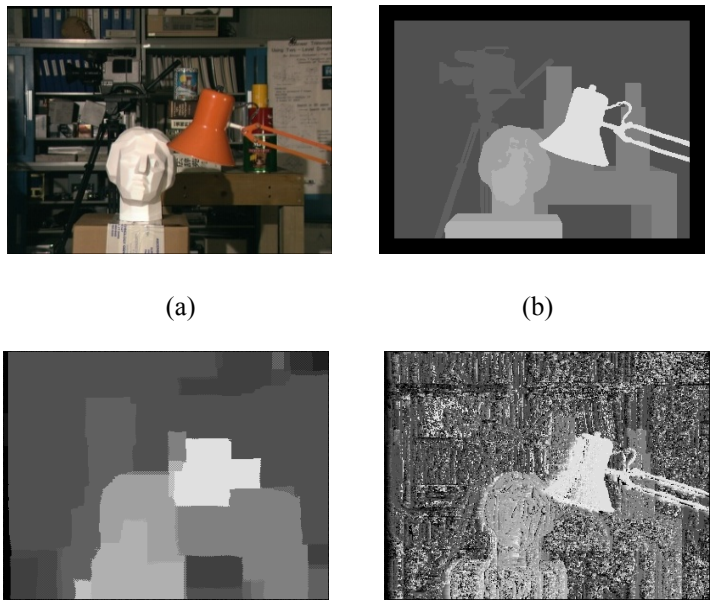
Say that p_L is a pixel in the left image that we want to match with its corresponding pixel, p_R, in the right image. Also, say that l_{p_L} is disparity (we will often call it disparity label or simply label) of p_L and q_L is a neighbor of p_L. In MRF algorithms, cost of matching p_L with p_R is defined by the data and smoothness terms coupled by several MRF parameters. In the following two subsections, we show how the MRF parameters influence the final labeling solution (i.e. the disparity map). Then we explain how we contribute in estimation of these parameters for better labeling solution than the existing algorithms'.

In MRF algorithm, matching cost of a site is estimated as the following

$$E_{p_L}(l_{p_L}) = D_{p_L}(l_{p_L}) + \lambda \sum_{q_L \in \mathcal{N}(p_L)} V(l_{p_L}, l_{q_L}), \quad (1)$$

where $D_{p_L}(l_{p_L})$ is the data cost (i.e., the likelihood term) of the point p_L at disparity l_{p_L}, $\sum_{q_L \in \mathcal{N}(p_L)} V(l_{p_L}, l_{q_L})$ is the smoothness cost (i.e., the prior term), $\mathcal{N}(p_L)$ is a set of neighbors of p_L, and λ is the smoothness parameter, which is also labeled as one of the MRF parameters. We will discuss formally more on the cost function of eq. (1) later introducing a few more parameters responsible for discontinuity handling. The objective of brief introduction here is to address the role of λ, which controls smoothness of the disparity solution, i.e., accuracy of the disparity map. As described by Figure 1.3, when λ is assigned a large number (=1000 for instance) the solution becomes over-smoothed; when assigned small (=0.01 for instance) the solution becomes under-smoothed. Thus, an appropriate value needs to be assigned to the parameter to obtain the best results.

Middlebury College provides a set of test images with known ground truth disparity map to evaluate the performance of stereo matching algorithms [Middlebury]. The algorithms ranked as the top performers according to the Middlebury College performance evaluation have good matching rates; however, in obtaining good results, the MRF parameters are heuristically determined or manually tuned. The problem with heuristic choice is that on an average they might work well for the test datasets but fail for unknown data. If the choice is manually tuned for test datasets with known ground truths, there is no way to verify if the parameters perform well on unknown data, since, knowledge on true depth of the surface is often unavailable. Recently, two automatic and statistical energy parameter estimation algorithms have been proposed – one by Zhang and Seitz in [Zhang05, Zhang07] and the other by Cheng and Caelli in [Cheng07]. These algorithms estimate the MRF parameters from the basic MRF expression of *maximum-a-posteriori* (introduced in Section 1.5). Both of these algorithms propose parameter estimation for existing MRF stereo matching algorithms using alternating optimization. These algorithms have two weaknesses. One of them is that while they provide good disparity solution overall, they do not perform well in handling discontinuity along the



(a) (b)

(c) $\lambda = 1000$ (d) $\lambda = 0.01$

Figure 1.3: (a) Reference stereo image, (b) ground truth disparity map, (c) over-smoothed solution, and (d) under-smoothed solution; the reference image and its ground truth disparity map are taken from Middlebury College test bed [Middlebury].

Surface borders. The other is computational costs involved in the estimation. In an alternating optimization, the optimization algorithm starts with an initial estimate of the parameters. Then, the parameters are applied to refine the disparity map and are updated (i.e., estimated) from the refined disparity map using *iterations*. The matching algorithm and parameter estimation algorithm continues alternatively until the parameters converge. Since the parameter estimation iterations are nested inside the matching iteration, computational complexities of these estimations are high.

1.3 Contributions

Contributions of this dissertation are three folds. First, it presents a novel MRF dense matching algorithm. A new cost function is modeled and a novel optimization technique relying on an adaptive support neighborhood mechanism is proposed for improved matching. We use homogeneous neighbors as the adaptive support neighborhood to handle discontinuity along surface borders and obtain better results than the existing algorithms' (see

Figure 1.4). Second, the proposed parameter estimation is computationally efficient. In the estimation, we update the data parameters *a priori* applying a noise equivalence hypothesis introduced by us. The smoothness parameters

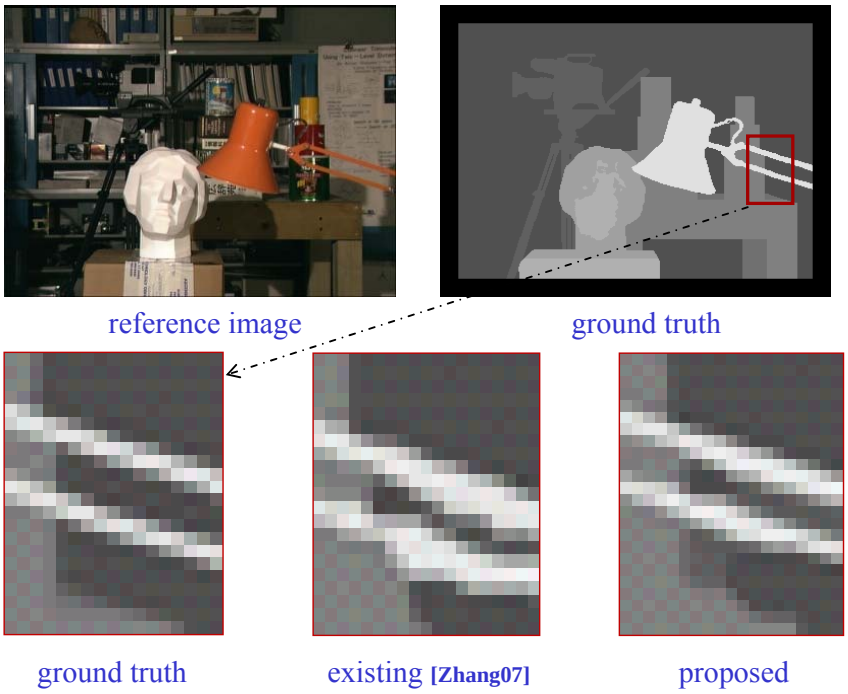


Figure 1.4: More consistent disparity map than existing algorithm with statistical estimation of parameters

are estimated applying a combination of maximum likelihood (hence, without using any iteration) and disparity gradient constraint. Therefore, the parameter estimation algorithms are computationally less expensive. Supporting experiments are carried out on ground truth datasets provided online by the Middlebury College [Middlebury].

In terms of computational costs, the differences of our parameter estimation approach compared to the two approaches in [Zhang05, Zhang07] and [Cheng07] are described in two generic flow diagrams in Figure 1.4. For more specific comparison, we explain computational complexities of the two approaches in Figure 1.5(a) and Figure 1.5(b) in number of operations. Let us say that each of the data and smoothness parameter estimation inference algorithms iterate a times on n number of pixels in S_L (the left image). Let us also say that the matching algorithm itself iterates b times on the n number of pixels in S_L. Then, according to Figure 1.5(a) the number of operations for estimation of both the data and smoothness parameters is $2abn$. According to Figure 1.5(b), the number of operations is $(a+b)n$. This calculation is for two-frame stereo, whereas in spac-etime stereo, the matching algorithm works on two videos of stereo images. Since we estimate the data parameters *a priori*, approach in Figure 1.5(b) spends computational time less than of approach in Figure 1.5(a). If the number of frames in each stereo video is c, approach in Figure 1.5(a) would require $2cabn$ operations, whereas, the proposed approach in Figure 1.5(b) would require only $(a+cb)n$ operations.

Finally, we propose an occlusion filling algorithm that fills the occlusions more consistently than the existing algorithms. Occlusions are regions in one of the stereo images that are not seen in the other stereo image. Figure 1.6 shows occlusion filling

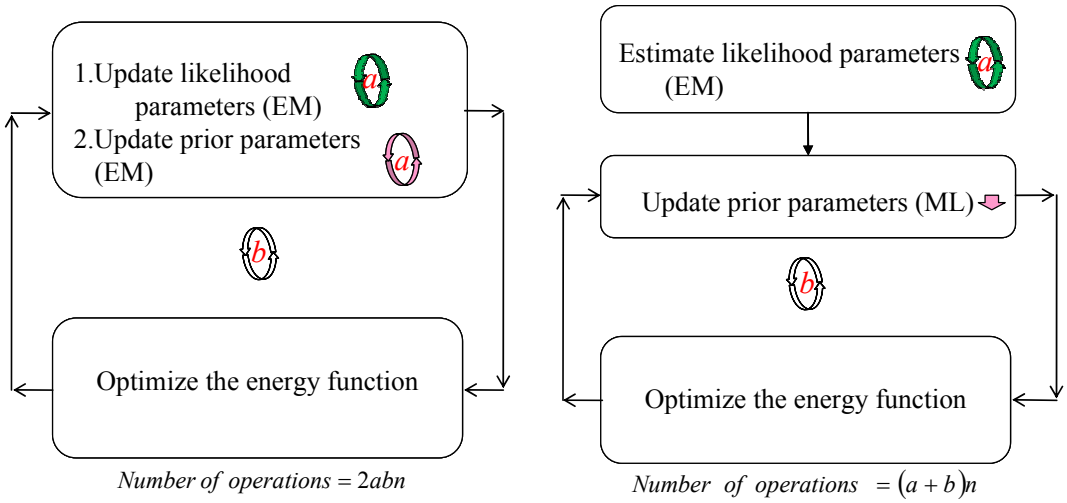


Figure 1.5: (a) Parameter estimation in existing algorithms [Zhang07, Cheng07] and (b) parameter estimation by the proposed algorithm

results of an existing and our proposed algorithms. Existing algorithms either fill occlusions without considering the slope of the occluded surface [Yang06c, Min08, Sun05] which is often similar to the slope of nearby non-occluded regions or they are not independents of the matching algorithm [Klaus06, Min08]. Occlusion filling from disparities of the neighbors is ambiguous when two backgrounds exist behind the foreground. In our proposed algorithm, we remove this ambiguity by proposing a measure called *homogeneity*, then apply interpolation, and achieve occlusion filling results with higher consistency.

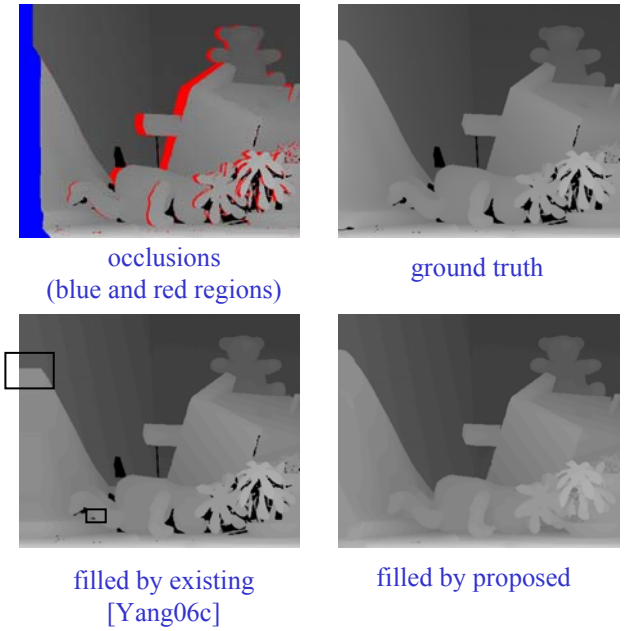


Figure 1.6: Occlusion filling results of existing and proposed algorithms; black rectangles show inconsistency in occlusion filling.

1.4 Motivations

For about the last four decades computer vision scientists have given a lot of effort to solve the stereo matching problem and improve the matching performance. While MRF stereo matching algorithms are the top performers among the existing stereo matching algorithms (according to the Middlebury College evaluation), the parameters in most of the algorithms are not automatic. In manual determination of the parameters, one has to visually verify that a particular parameter setting produces more correct matches than the other settings – a difficult task without any knowledge of true structure of the scene. When the parameters are determined empirically from ground truth data sets, there is always a possibility that the parameters will fail to generate correct matching results for datasets providing no knowledge on their surface topologies.

Our research in stereo vision in general is motivated by the capabilities the stereo vision technique offers in solving many real life problems. In applications that require 3D reconstructions, measurements, or navigation without being too close to the objects, stereo vision is a suitable technique [Huq07a, Kittler05, Onofrio05, Huq04, Chen01]. For instance, stereo vision has successfully modeled 3D human faces for face recognition allowing the recognition to happen without requiring any physical contact. 3D maps of earth surfaces with natural objects, such as, mountains, rivers, etc. or human made objects, such as, roads, buildings, etc. are often the outcomes of the satellite stereo images. Stereo images captured with microscopes facilitate nano-scale 3D surface metrology and inspection for fracture, corrosion, or accumulation of elements of interests [Huq08a, Huq07b, Kammerud05, Hao07, Huq06]. In surgery, doctors can determine pose of a patient with the help of stereo vision technique and take step accordingly; a vehicle can estimate distances of obstacles in the streets and highways to avoid accidents or to drive autonomously. In general, stereo vision contributes in quality control with surface inspection and measurements, in safety and security with remote sensing and navigation, and in reverse engineering where 3D CAD models are recovered from objects of unknown dimensions. Some real life applications are illustrated with pictures in Figure 1.7. In these applications, parameters in the matching algorithms need to be estimated automatically to adapt with varying imaging conditions and scene structures and obtain reliable disparity solution.

In many navigation applications, intelligent robots may need to map terrains for path planning and localize objects (often in real-time). To visualize the surrounding environments into 3D, the robots can rely on stereo vision systems. Performance of the stereo matching algorithms in terms of matching error and runtime are crucial to make inferences properly from the 3D maps of the surroundings captured by these robots. The performance is related to the work presented in this dissertation.

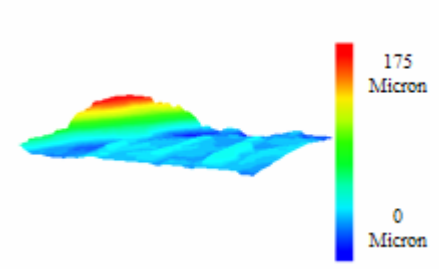
1.5 The cost function and Bayesian inference

In this section, we discuss how the cost function in stereo matching, which is comprised of the data and smoothness terms, is derived from Bayesian inference [Rosenfeld76]. In Bayesian inference, if an object is given and we want to find its class, we select a class (often a particular distribution) first and then find the probability of that object of falling into that class. We assign the object the class with the highest probability. Similarly, in stereo matching, we want to classify each pixel into their corresponding class of labels (the disparities). According to the Bayesian inference, we assign a label to a pixel and then estimate the probability of that pixel having that label. We pick the label for which the probability is the highest. Following is a mathematical interpretation of this inference.

Say that $|S_L| = m$ is the total number of pixels in S_L, $p_L \in S_L$ is a pixel in the left image, and $\Lambda = \{w_1, w_2, ..., w_k\}$ is the set of allowable labels to be assigned to each pixel, p_L. Also say that $\mathcal{L} = \{l_1, l_2, ..., l_m\}$, $l_j \in \Lambda$ is a combination of labels, which maps between S_L and Λ, i.e. $\mathcal{L} : S_L \to \Lambda$. Say Ω is the space of $\mathcal{L}$ (then, $|\Omega| = |\Lambda|^m = k^m$). Our goal is to find an estimate, $\mathcal{L}^*$, such that the following,

$$\mathcal{L}^* = \underset{\mathcal{L} \in \Omega}{arg\max}\, P(\mathcal{L} \mid S_L) \propto \underset{L \in \Omega}{arg\max}\, P(S_L \mid \mathcal{L}) P(\mathcal{L}) \tag{2}$$

3D Face Modeling
(IRIS)



Surface inspection in micro-scale
(IRIS)

3D Urban Modeling
(University of North Carolina)

Unmanned Vehicle
(Seibersdorf Resesarch)

Figure 1.7: Some applications of 3D stereo modeling

is satisfied in accordance with the Bayesian inference. Equation (2) can be expanded using the joint probability of $\mathcal{L}$,

$$\mathcal{L}^* \propto \underset{\mathcal{L}\in\Omega}{arg\max}\, P(S_L \mid \mathcal{L})P(\mathcal{L}) = \underset{\mathcal{L}\in\Omega}{arg\max}\left\{ \prod_{p_L\in S_L} P\left(x_{p_L} \mid l_{p_L}\right)\right\} \bullet P(\mathcal{L}),$$

where x_{p_L} is an evidence at the site p_L; x_{p_L} could be, for instance, an absolute intensity difference of the pixel p_L and its corresponding pixel p_R for a label l_{p_L}. The domain of $\mathcal{L}$, which is Ω, is large (recall $|\Omega| = |\Lambda|^m$). Ω turns into a small domain when the following two assumptions are applied [Besag86]:

1. Markov Random Network assumption on $\mathcal{L}$, which means that labeling of a pixel is conditionally dependent only on its neighbors' labels.

II. Each pixel has the same conditional density function on its label and they are independent (we can relax this assumption by allowing conditional density function to be different for each pixel).

Applying the combination of assumption I and Hammersley-Clifford theorem (1971), eq. (2) can be reduced to an optimization problem with conditional probability. Hammersley-Clifford theorem says, if $\mathcal{L}$ is a Markov Random Field, assumption II will imply,

$$P(\mathcal{L}) = \prod_{p_L\in S_L} P\left(l_{p_L} \mid \left\{l_{q_L} : q_L \in \mathcal{N}\{p_L\}\right\}\right),$$

where $\mathcal{N}\{p_L\}$ is a set of neighbors of p_L. Then, by conditional probability theorem,

$$\mathcal{L}^* \propto \underset{\mathcal{L}\in\Omega}{arg\max}\left\{ \prod_{p_L\in S_L} P\left(x_{p_L} \mid l_{p_L}\right)\right\} \bullet P(\mathcal{L}) = \underset{\mathcal{L}\in\Omega}{arg\max}\left\{ \prod_{p_L\in S_L} P\left(x_{p_L} \mid l_{p_L}\right)P\left(l_{p_L} \mid \left\{l_{q_L} : q_L \in \mathcal{N}(p_L)\right\}\right)\right\} \quad (3)$$

The first probability term in eq. (3) is the likelihood of the site p_L and the second term is its prior (spatial coherence or smoothness). x_{p_L} is an evidence for p_L at a given label l_{p_L}. Both terms together is the *maximum-a-posteriori* (MAP) of p_L, a_{p_L}, written as

$$a_{p_L} = P\left(x_{p_L} \mid l_{p_L}\right)P\left(l_{p_L} \mid \left\{l_{q_L} : q_L \in \mathcal{N}(p_L)\right\}\right). \quad (4)$$

$-\log a_{p_L}$ is equivalent to a constant plus summation of the exponent terms of the likelihood and prior. Therefore, optimizing (4) equals to minimizing summation of the negative of the exponents of both the likelihood and prior terms defined as energy of the pixel p_L, $E_{p_L}\left(l_{p_L}\right)$, for a label l_{p_L}. If $D_{p_L}\left(l_{p_L}\right)$ is the negative exponent part of the likelihood and $V\left(l_{p_L}, l_{q_L}\right)$ is of the prior, then

$$E_{p_L}(l_{p_L}) = \lambda_1 D_{p_L}(l_{p_L}) + \lambda_2 \sum_{q_L \in \mathcal{N}(p_L)} V(l_{p_L}, l_{q_L}), \tag{5}$$

where λ_1 is the likelihood (alternatively the data) and λ_2 is the prior (alternatively the smoothness) model parameters. Since we want to optimize $E_{p_L}(l_{p_L})$, estimation of absolute energy is not necessary. Consequently, we can eliminate one of the parameters and write $E_{p_L}(l_{p_L})$ as

$$E_{p_L}(l_{p_L}) = D_{p_L}(l_{p_L}) + \lambda \sum_{q_L \in \mathcal{N}(p_L)} V(l_{p_L}, l_{q_L}). \tag{6}$$

The energy equations (5) and (6) have been used in stereo vision literature quite often for solving stereo matching iteratively with variations in the definition and estimation of the λ-parameters and optimization of $E_{p_L}(l_{p_L})$, thereby with varying performances. If we drop the assumption II partially, i.e., allow each site to have the same density function on the likelihood and varying but independent conditional density function on their priors, eq. (6) can be rewritten as eq. (7).

$$E_{p_L}(l_{p_L}) = D_{p_L}(l_{p_L}) + \sum_{q_L \in \mathcal{N}(p_L)} \lambda_{p_L q_L} V_{p_L q_L}(l_{p_L}, l_{q_L}). \tag{7}$$

In eq. (7), the smoothness term $V_{p_L q_L}(l_{p_L}, l_{q_L})$ and the parameter $\lambda_{p_L q_L}$ are now dependent on a neighbor the site p_L interacts with. Nonetheless, densities of the priors, i.e., interactions of p_L with different neighbors, q_L, are still independents.

The concept of MRF for labeling with the estimation of MAP in eq. (4) was first introduced by Rosenfeld et al. in 1976 [Rosenfeld76]. Algorithms that apply this concept are called relaxation labeling algorithms. Thus, relaxation labeling is defined as a class of iterative algorithms for combinatorial optimization that assigns each site a label taken from a set of discrete labels. As the iteration continues, ambiguity in labeling is reduced. Relaxation applied with eq. (4) is a probabilistic relaxation. On the other hand, relaxation with eq. (6) (or (7)) is energy minimization-based. In our case, the both kinds are equivalent since the later ones are derived from the former. Relaxation labeling has been widely used in many image processing and computer vision applications such as image enhancement, edge detection, motion detection, image segmentation, object matching, and so on.

Our goal is to minimize $\sum_{p_L \in S_L} E_{p_L}(l_{p_L})$ for obtaining an optimum configuration $\mathcal{L}^*$. $-E_{p_L}(l_{p_L})$ is not guaranteed to be a convex function, since, even if the smoothness term is chosen to be convex, the data term is never guaranteed to be convex. Note that eq. (6) (or (7)) is not discontinuity preserving either and needs modification for it to have such ability. Discontinuity occurs at surface borders where disparity (label) has sudden jumps (increase or decrease). If these jumps occur along horizontal direction, portions of the image near the boundary, which are visible in one of the images, are invisible in the

other. These regions that become invisible in the other image are called occlusions. It is because of the discontinuity handling and occlusion detection stereo matching has become even a harder problem to solve.

1.6 Modeling and optimization in MRF stereo matching

Before going further into details in the remaining chapters of this dissertation this section provides a quick overview on the modeling and optimization strategies of the cost functions in relaxation labeling. The purpose is to make the remainder of this dissertation more readable to the readers.

We model the cost function first and then optimize as described by the block diagram in Figure 1.8. In the modeling phase, we model the data term from likelihood, which is a data cost that follows a probability distribution or we model the data term with other data costs such as sum of absolute differences or cross correlation coefficient. For likelihood, the probability distribution and its parameters could be selected based on assumption or based on evidences from the ground truth data. Because of the ambiguity in matching, contextual constraints such as spatial coherence is introduced in the modeling. Spatial coherence basically says that labeling of neighboring pixels is smooth, i.e., two neighboring pixels should have similar labels unless they are truly discontinuous, i.e., they are on two adjacent surfaces (in the image space) with relatively big depth differences (in the object space). Handling discontinuity could involve introducing more parameters into the modeling phase. Spatial coherence (smoothness) between p_L and its neighbor is defined with an interaction function, which could also follow a distribution. The interaction function is again modeled based on assumptions or based on evidences from the ground truth data. To enforce the spatial coherence strongly, p_L interacts simultaneously with multiple neighbors that form a neighborhood system of an order. A few neighborhood systems of different orders are illustrated in Figure 1.9. For a particular algorithm, often, one of these systems is chosen empirically.

Input data could have more priors. For instance, another prior in stereo is visibility constraint that says that corresponding pixels in the right image shift only in the left direction (see Figure 1.1 (b)). The visibility constraint is true for all the pixels and can be enforced in implementation of the matching algorithm explicitly; therefore, this prior is not necessary to be included in the cost function.

The second phase is optimization, which is stochastic because of the nature of the stereo matching problem and iterative because of its origination from the Bayesian inference. To obtain a global minimum of $\sum_{p_L \in S_L} E_{p_L}\left(l_{p_L}\right)$, one needs to introduce a global convergence mechanism either *implicitly* or *explicitly*. Apart from the model parameters of the data and smoothness costs (i.e., the likelihood and prior terms respectively in the probabilistic framework), other parameters that control the global convergence mechanism explicitly could be introduced in eq. (6). Nonetheless, it is worth mentioning the theoretical fact that, because of having a stochastic nature, no optimization algorithm

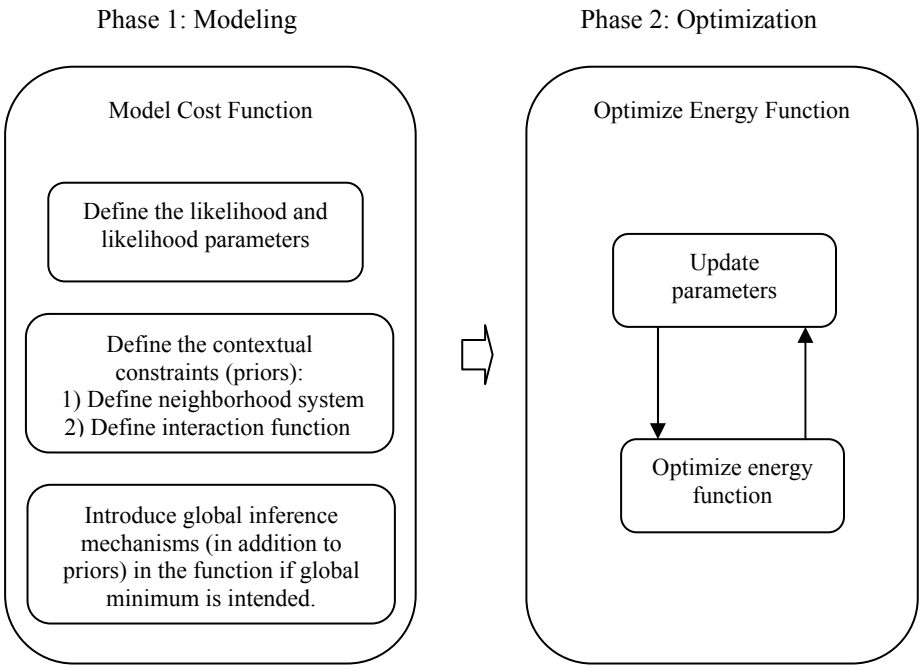


Figure 1.8: Modeling and optimization strategies diagram

for stereo matching could guarantee a global solution to the optimization of $\sum_{p_L \in S_L} E_{p_L}(l_{p_L})$.

Since the parameters are dependent on the labels (the disparities) of the pixels, an alternating optimization algorithm is adopted (except for the data parameters in this dissertation, which have been estimated *a priori*). In an alternating optimization approach, the parameters are updated from available current labels of the pixels. Then, updated parameters are used for a refined labeling through optimization of the aggregated cost of the sites. The alternating phases continue until the aggregated cost converges.

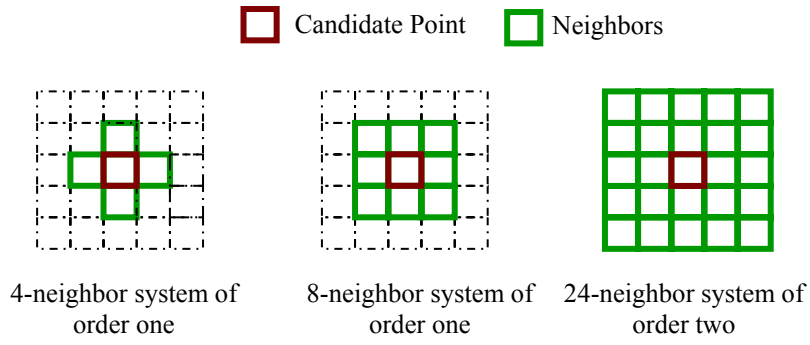


Figure 1.9: Neighborhood systems of different orders

1.7 A block diagram of our approach

Figure 1.10 presents a block diagram of the proposed MRF stereo matching algorithm along with occlusion filling to obtain an occlusion filled disparity map of a scene from binocular stereo images. As described in the diagram, first we perform the stereo matching. In stereo matching, the data model parameters are estimated *a priori* from a single image and adaptive support neighborhood selection is performed before the matching iteration starts. In the matching iteration, we perform the stochastic optimization alternated with the estimation of the smoothness parameters. The smoothness parameter is estimated using maximum likelihood, hence, only a single iteration. The matching iteration works on a pair of symmetric functions. One of these symmetric functions performs left-to-right matching and the other performs right-to-left. The *a priori* estimation of the data model parameters provides stability in convergence of the optimization step. This is because the data and smoothness parameters are interdependent; one of them can not be known unless the other one is already known. The *a priori* estimation breaks interdependence in estimation of the two parameters. The adaptive support neighborhood adds additional consistency in handling of the discontinuity in smoothness. In matching, we enforce a number of constraints and assumptions. They are discussed in details in the next subsection. After the matching converges, occlusions are detected by comparing the two disparity maps delivered by our symmetric matching algorithm and then, occlusions are filled.

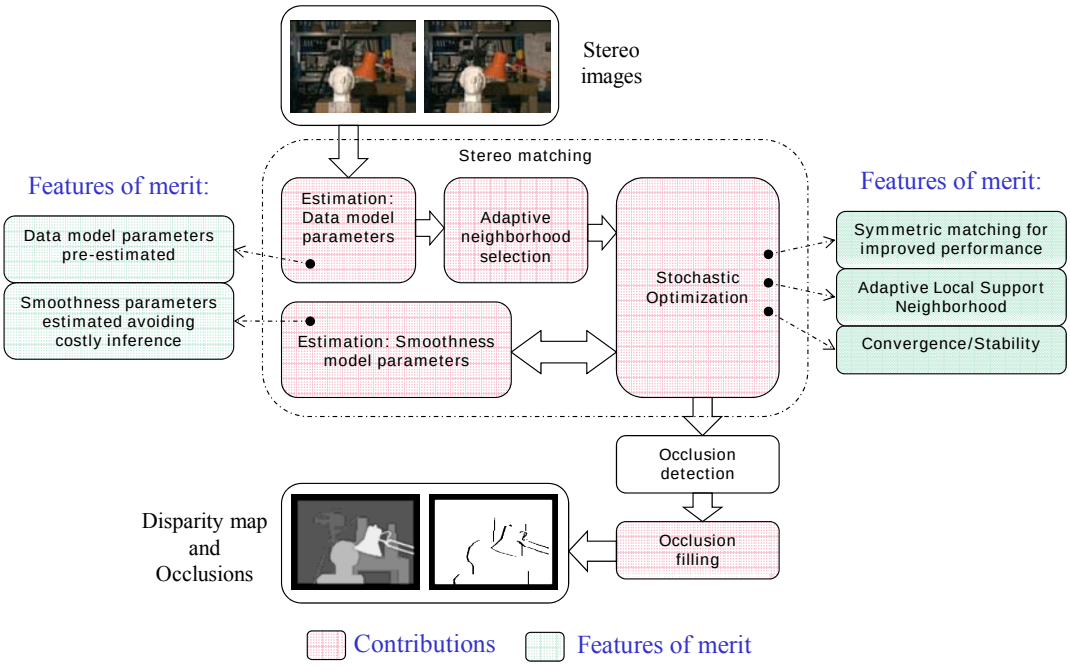


Figure 1.10: The flow diagram of our proposed MRF stereo matching algorithm and occlusion filling

1.8 Constraints and assumptions in matching

Stereo images could have several constraints and a number of assumptions can be imposed on them. These constraints and assumptions can be effectively applied into stereo matching algorithms to achieve good matching results. As mentioned earlier, one of the constraints is the visibility constraint: any corresponding pixel in the right image will shift to the left. One of the assumptions is rectification. All the matching algorithms proposed or cited in this dissertation assume that the stereo images are rectified. When a scene is captured into two images from two viewpoints, corresponding image points, which are projection of the same scene point, may not appear on the same scan line of the stereo images. From camera epipolar geometry (Figure 1.11(a)) it is possible to transform the images such that the corresponding image points appear on the same scan line [Fusiello00, Trucco98] in both images and also, all the points on a left scan line appear on the corresponding right scan line. Figure 1.11(a) describes the epipolar geometry where C_L and C_R are the camera centers and e_L and e_R are the epipoles. On the image planes, p_L and p_R are the projections of the same scene point P. An epipole of a camera is the projection of the center of the camera on the image plane of the other camera. All points on an epipolar line, which connects the epipole with the projection of a scene point, in one image plane lies on the corresponding epipolar line in the other image plane. From a set of known pair of matched points it is possible to align the points on the corresponding epipolar lines horizontally on the same scan lines in both images as described by Figure 1.11(b). This transformation process is called rectification. For more on the epipolar geometry and rectification we refer to [Faugeras93, Hartley04]. Since we assume the images to be rectified, search for correspondence in the right image is not necessary in other directions except along the horizontal line.

Another assumption we make is that the object surfaces are Lambertian [Lambert60], which means that brightness of a point does not change if viewed from another angle. We hold this assumption for both kinds of images – images from the optical cameras and images from the LC-SEM (Large Chamber Secondary Electron Microscope). This assumption leaves modeling of the image noise to take into account only the other sources such as dark current, discretization, etc. We further assume that all noise, generated from these sources, combined follows a single distribution.

There are three other constraints – disparity gradient, ordering, and uniqueness. Disparity gradient constraint says that an object surface can not be bent more than a threshold, called disparity gradient threshold. For narrow baseline stereo this threshold value is small and often assumed as 1.0. We hold this assumption in our proposed stereo matching algorithm. Violation of disparity gradient constraint occurs due to depth discontinuity, which, when occurs along the horizontal line, leads essentially to occlusions. On the other hand, ordering constraint says that consecutive sites in the left image appear in the same order in the right image. This constraint is not valid if narrow objects exist in the scene and they are located relatively close to the camera with respect to their backgrounds. We do not apply ordering constraint, but we apply the uniqueness constraint that allows a site in one of the images to have only one match in the other.

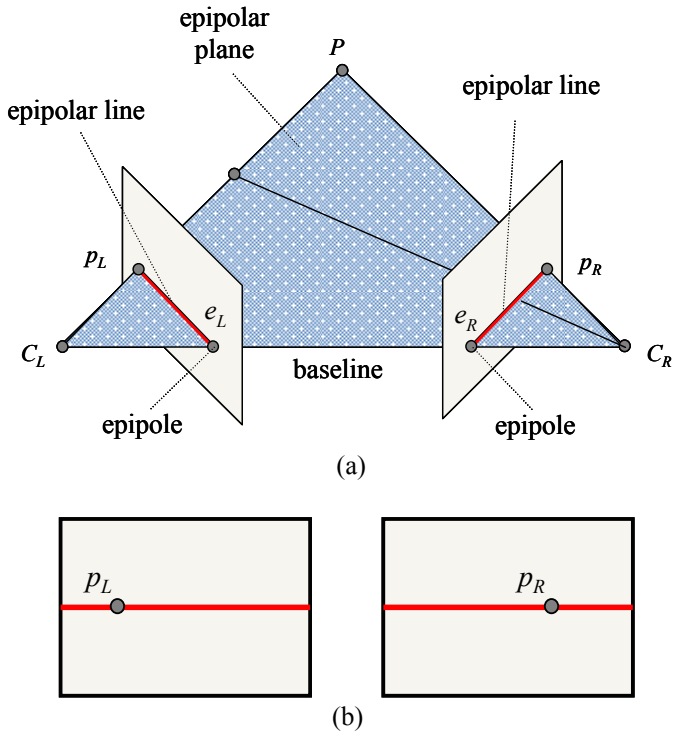


Figure 1.11: (a) epipolar geometry and (b) rectified images

1.9 Organization of the dissertation

In chapter two, we review the existing energy minimizing (MRF and non-MRF) stereo matching algorithms. States of the art of stereo matching are presented with classification based on their optimization techniques and with performance in terms of matching error and parameter estimation techniques. We also present a survey summary of state of the art in occlusion filling. In chapter three, we describe modeling of the proposed cost function in details, introduce the MRF parameters and their statistical estimations, describe the adaptive support neighborhood selection algorithm, and describe the optimization of the aggregated cost function. We propose symmetric matching cost functions for improved matching and occlusion detection. In chapter four, we present a number of existing and proposed occlusion filling algorithms. We detail the occlusion filling problem that has not been systematically defined and studied in the current stereo vision literature.

In chapter five, we present experimental results. We conduct experiments on the Middlebury College datasets to show that with the proposed parameter estimation our adaptive support neighborhood based matching algorithm performs with matching error lower than the existing algorithms'. The Middlebury College datasets are available online with their ground truth matching information for each pixel of the stereo images and are used by authors worldwide to evaluate performances of stereo matching algorithms. We

also present matching results from applying the proposed algorithm on LC-SEM images. We apply existing and proposed occlusion filling algorithms to the ground truth disparity maps with occlusions labeled. The best proposed occlusion filling algorithm is then used to fill occlusions in the disparity maps generated by our matching algorithm. The occlusion filled disparity maps are compared both numerically and visually with existing algorithms that work with manual settings of the parameters. Finally, we implement our parameter estimation technique with an existing top ranked algorithm called BP (belief propagation) and show that our estimation technique generates comparable results with computational time less than the existing BP algorithms'.

Chapter six draws conclusions and suggests a number of future works based on the novel ideas and concepts applied in our proposed algorithms. An appendix, which is related to the subject of this proposal, is added at the end. We develop a generic method for calibration of SEM/LC-SEM scanners assuming an affine camera model. This calibration model is for a more generalized image capturing set up than the ones existing in the literature; the existing models are derived from our generalized model.

2 Literature review

Research in stereo matching has continued for almost the last four decades. Matching algorithms, developed in the early period, mostly focused on sparse matching at salient image features, such as, corners and edges. Taking advantage of the growing computing power, lately computer vision scientists have focused on dense matching, where, images are matched at every pixel. Since the beginning of the current decade, dense matching has drawn significant attention. The top performers of these algorithms are cost function (i.e. energy function) minimization based and they are reviewed in this chapter. First we review the algorithms in terms of their optimization techniques. Stan Z. Li has classified generic MRF optimization techniques for computer vision in his book [Li95]. Our classification in this chapter is specifically for MRF stereo. Then, we focus on existing MRF stereo matching algorithms with statistical parameter estimation. The algorithms are studied comparatively based on their performances on the Middlebury College ground truth datasets. Finally, we conduct survey on a few existing stereo matching algorithms to project upon the states of the art of techniques for occlusion filling undertaken by those algorithms.

2.1 MRF stereo matching algorithms

Performances of an MRF stereo matching algorithm depend on several main factors: modeling of the cost function, the MRF parameters and their estimations, and optimization of the cost function. As mentioned earlier, the optimization technique is generally stochastic because of the stochastic nature of a stereo matching cost function. Even if the potential function (the smoothness cost) is convex (i.e., in L_2 or higher norm), eq. (6) is considered as a stochastic function due to the fact that in an allowable range of l_{p_L}, the data cost function is stochastic. Therefore, optimization of $\sum_{p_L \in S_L} E_{p_L}(l_{p_L})$ for a global minimum is stochastic. Because of being stochastic, there are two important facts with all existing MRF stereo matching algorithms. One, existing methods do not guarantee a global minimum. Two, starting with a different initial labeling (i.e. disparity map) in different runs these algorithms do not guarantee the same final labeling.

One of the bases for classification of the MRF stereo matching algorithms is then how their underlying cost functions are optimized. On this basis, we can study them as local methods and global methods, which are also the major classifications of the stochastic

optimization techniques. Several algorithms have aspects both local and global in nature. They are reviewed as another class called 'Hybrid methods'. Besides, one class of algorithms, which is global in nature, has been developed specifically for solving stereo matching problem. This class is known to the vision community as dynamic programming. Earliest to most recent algorithms are divided in these four classes for review in the following subsections.

2.1.1 Local methods

In the local methods, matching is performed by taking one site at a time. Iterated conditional mode (ICM) is a local method well known for being one of the earliest stochastic optimization techniques. ICM is briefly described here.

ICM (Besag 1986 [Besag86]): In ICM, the cost function in eq. (6) is optimized individually at every site, but the sites are accessed in a certain order. The optimization technique itself was first proposed by Besag [Besag86]. Because of having an access direction, ICM is labeled as a deterministic approach sometimes. In stereo matching case, one possible direction is to pick the site that has the smallest row and column number among the sites that have not yet matched in the current iteration. Such an order ensures that the matching information of one site is utilized by (hence propagated to) an immediate neighbor. Since the sites are optimized one at a time, ICM provides only a local solution. ICM has been used for comparing with other techniques, such as in [Szeliski06]. Shafik et al. have used ICM for 3D face modeling from stereo images [Huq07].

Another commonly used local method is Winner Take All (WTA) where the stereo matching is performed on all allowable labels of a site. Label that requires the minimum energy is selected. [Huq08b] is an example work of WTA stereo matching algorithm. WTA has often been called Highest Confidence First (HCF).

2.1.2 Global methods

In the global methods, an aggregated cost over all the sites in the left image is optimized. Variation in aggregated cost occurs through annealing of the MRF parameters common to all the sites. In global optimization methods, eq. (6) can be rewritten as an aggregated energy,

$$\sum_{p_L \in S_L} E_{p_L}(l_{p_L}) = \sum_{p_L \in S_L} \left\{ D_{p_L}(l_{p_L}) + \lambda \sum_{q_L \in \mathcal{N}(p_L)} V(l_{p_L}, l_{q_L}) \right\}, \tag{8}$$

where λ is the smoothness parameter common to all the sites. Equation (8) is again not guaranteed to be a convex function. Four techniques for the global optimization of eq. (8) are discussed here.

Simulated annealing (Kirkpatrick et al. 1983 [Kirkpatrick83]): First introduced by Kirkpatrick et al. in 1983, simulated annealing was used in computer vision for image restoration with *maximum-a-posteriori* (MAP) estimation by Geman and Geman [Geman84] and later for solving stereo matching problem by Barnard [Barnard89]. Simulated annealing is a common cost minimizing technique, where the annealing parameters (i.e. the MRF parameters in stereo) are varied within a range the cost function is believed to have the global minimum. Even though finding a range is often easy, difficulty arises in determining the step size for increment of a parameter for a possible hit to the minimum – for a big step size the minimum could be missed, but for a small size computational time goes up exponentially with an increase in the number of parameters to be varied. In stereo matching, the smoothness parameter could be common to all the sites as shown in eq. (8) or vary site to site (depends on the modeling assumptions of the cost function) as shown in eq. (7). If common, quality in matching is sacrificed; if varied to achieve the quality, computational time becomes unacceptably high. Assuming a common smoothness parameter, simulated annealing was used in [Szeliski06] to study the relative performances of other MRF stereo matching algorithms.

Mean field annealing (Peterson and Soderberg 1989 [Peterson89], Geiger and Girosi 1989 [Geiger89]): Mean field annealing was first introduced by Peterson and Soderberg in 1989 and independently by Geiger and Girosi in the same year. Mean field annealing methods replace the stochastic update rules of simulated annealing with deterministic rules based on the behavior of the mean disparity at each pixel [Geiger91]. Nonlinear diffusion used in [Scharstein98] for solving stereo matching problem is a mean field annealing algorithm; occlusions in this algorithm are not detected but treated only as noise.

Graduated non-convexity (Blake and Zisserman 1987 [Blake87]): Developed by Blake and Zisserman, graduated non-convexity intends to reach the global optimum starting with a convex version of eq. (8) and ending at the desired non-convex function in eq. (8). At each step, non-convexity is adopted by choosing another convex version of eq. (8). The difficulty of this algorithm lies in finding a non-convexity rate and related cost functions which gradually turn into the desired function. Graduated non-convexity was applied by Oriot and Besnerais for matching aerial stereo images [Oriot98]. The data term was window based cross-correlation coefficient. Gradual non-convexity was applied only to the smoothness term, which, however, did not guarantee that both terms together guaranteed convexity, hence, still had a chance to be trapped at a local minimum. In their algorithm, disparity gradient was varied from a predefined maximum to a minimum (=1.0) to apply non-convexity.

Genetic algorithm (Holland 1975 [Holland75]): Genetic algorithm (GA) was first proposed by John Holland [Holland75]. GA is a general purpose global optimization technique, which is iterative, randomized search based, and often regarded as an alternative method for solving complex optimization problems, especially for functions whose derivatives can not be computed numerically.

In GA, first, an initial population of solutions is generated. Since the population is huge (for stereo matching, $|\Omega| = |\Lambda|^m$), usually, cross-correlations with multiple windows of different sizes are applied first to reduce the number of allowable labels for each site. Then, all sites are put together in a string called chromosome, where one label among the allowable set of labels for each site is a gene. The population then goes through crossover and mutation operations to generate a new population from where a new generation of population is created by estimating fitness of each chromosome and picking only those having high fitness scores. The crossover and mutation operations assign a site a new label (i.e. a disparity value) making sure that the uniqueness constraint is maintained. Ordering constraint is often enforced during these operations.

GA was applied in [Gong01, Saito95, Tien04] for stereo matching. For dense matching in [Gong01], the motivation was to improve the accuracy of the disparity map generated after removing the mismatches caused by both occlusions and false matches. The algorithm took advantage of multi-view stereo images to detect occlusions, therefore, removed mismatches caused by the visibility constraint. In [Tien04], the matching was sparse and performed only on edges. Both in [Gong01] and [Tien04], chromosome fitness function for selection with elitist strategy [Holland75] was similar to eq. (6). In [Saito95], which was also a dense matching algorithm, fitness function was only the absolute difference of intensities. Besides, the reference image was divided into blocks and GA was applied to each block separately. The chromosome length was equal to the number of sites in one block.

GA has not been used much probably because stereo matching is only a difficult but not a complex problem. As mentioned, GA is better suited to solve complex problems where parameters can not be estimated explicitly.

2.1.3 Hybrid methods

Hybrid methods deploy partially global and local aspects in their optimizations. We have described here two hybrid methods; these two methods and their variants are currently the top performers in solving stereo matching problem. The first method is graph cuts. In graph cuts, one iteration i is for one of the allowable labels, w_i, to check against all the sites, which we consider as a global aspect. If the aggregated cost in eq. (8) estimated in the current iteration is smaller than in the previous iteration, sites that do not have the label w_i is now labeled as w_i. On the other hand, the checking on whether the cost of assigning the new label w_i to a site is lower than its current label is performed locally at that site. The second method is belief propagation, which has local cost computation at each site but also has message passing between neighboring sites. We consider this message passing, which can be thought of information sharing among sites within a large neighborhood, a global (or semi-global) aspect. Graph cuts and belief propagation are described in more details here.

Graph cuts (Menger 1927 [Menger27], Elias et al. 1956 [Elias56], Ford and Fulkerson 1956 [Ford56]): Graph cuts as a combinatorial optimization technique was first proposed and applied to stereo matching by Boykov et al. in 2001 [Boykov01]. The central idea was to represent an energy function (eq. (8)) with graphs and then optimize the function through minimum cuts. Theory of minimum cuts of a graph was developed long ago by Karl Menger in a graph theory in 1927 [Menger27]. This theory says that if *s* and *t* are distinct, non-adjacent vertices in a connected graph *G*, then the maximum number of *s*-*t* paths in *G* equals the minimum number of vertices needed to separate *s* and *t*. This theory derives the max-flow min-cut theory: the value of the max flow is equal to the value of the min cut, which was proved by P. Elias, A. Feinstein, and C. E. Shanon in 1956 [Shanon56], and also independently by L. R. Ford and D. R. Fulkerson in the same year [Ford56].

Graph cuts operates on graph-representable functions of binary variables. In stereo matching, each site can have more than two labels hence they are not binary. To treat the sites as binary variables only, in a matching iteration two choices are given to all the sites: either the site moves to the label w_i or it remains with whatever label it has. To be graph-representable, a function has to be regular. A function is regular if it meets the following constraint,

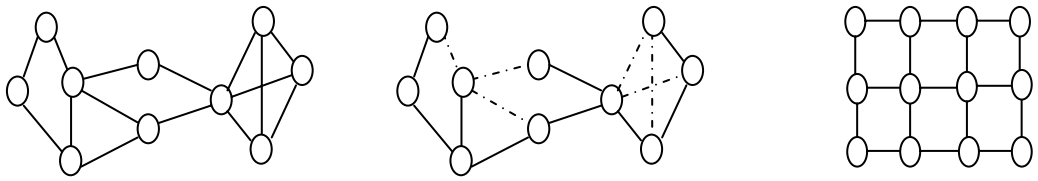
$$E(0,0)+E(1,1)\le E(0,1)+E(1,0), \tag{9}$$

where $E:\{0,1\}^2\to\mathcal{R}$ is a function of the 2nd order cliques of binary variables. For $E_{p_L}(l_{p_L})$ to be a regular function only $V(l_{p_L},l_{q_L})$ needs to be a metric since $D_{p_L}(l_{p_L})$ is a unary function, hence, trivially regular. $V(l_{p_L},l_{q_L})$ is regular when it is a metric; since then, the triangular relation holds for the constraint in (9) to be true. Graph cuts uses $D_{p_L}(l_{p_L})$ as a quadratic of absolute difference between intensities of p_L and its corresponding site p_R at l_{p_L} and $V(l_{p_L},l_{q_L})$ as a truncated Euclidean distance, where Euclidian part makes $V(l_{p_L},l_{q_L})$ metric and truncation makes it discontinuity preserving. Thus,

$$V(l_{p_L},l_{q_L})=\min\left(K,\left|l_{p_L}-l_{q_L}\right|\right), \tag{10}$$

where K is a free parameter working as a bound on $V(l_{p_L},l_{q_L})$. Another choice for $V(l_{p_L},l_{q_L})$ in graph cuts is Potts interaction penalty [Potts52], which is also both metric and discontinuity preserving (eq. (11)).

$$V(l_{p_L},l_{q_L})=\delta\left(l_{p_L}\neq l_{q_L}\right), \text{ where } \delta(true)=1 \text{ and } \delta(false)=0. \tag{11}$$



(a) (b) I

Figure 2.1: (a) Graph with a loop, (b) graph without loop, and (c) loopy graph created by a 3×4 stereo image; each matching site is a node in the graph.

Both (10) and (11) have demerits. Equation (10) has an unknown parameter K to be estimated. Equation (11), on the other hand, sacrifices matching quality by replacing slanted surfaces with stair like structures.

Graph cuts based existing algorithms [Hong04, Yang06a] are good at preserving discontinuity (occlusion or discontinuity of surface smoothness). However, they are sensitive to noise, since the data and smoothness terms the algorithms work with are not statistical rather chosen heuristically. The smoothness term described by (11) imposes piecewise linearity on surfaces, hence, leaves out stair effects in reconstructions. In an experiment by Szeliski et al., it is observed that at best, Graph cuts produces aggregated cost that is 0.018% over the global minimum (obtained from the ground truth data), while at worst, the cost is 3.6% larger [Szeliski06].

Belief propagation (Pearl 1986 [Pearl86], Lauritzen and Spiegelhalter 1986 [Lauritzen86]): Judea Pearl in 1986 and Lauritzen and Spiegelhalter in the same year independently formulated belief propagation. Belief propagation (BP) works on finite cycle-free graphs. However, because all its operations are local, it may also be applied to graphs with loops (hence calling it as loopy belief propagation) (see Figure 2.1), such as stereo images, where it becomes iterative and approximate. In BP, each site estimates the MAP and passes the MAP information, called message, to its neighbors.

There are two versions of BP algorithm, sum-product and max-product; both the algorithms produce the same results [Felzenswalb04]. We discuss here only the later one since it resembles to one of the cost functions (eq. (6)) described in this dissertation. The max-product BP algorithm works by passing messages around the graph defined by the 4-neighbor systems of order one (see Figure 1.9 and Figure 2.2). Each message is a vector of dimension given by the number of possible labels. Let $m^t_{q_L p_L}$ be the message that node q_L sends to a neighboring node p_L at time *t*. s_L are all the neighbors of p_L except q_L, the neighbor to whom message is sent. All entries in $m^0_{q_L p_L}$ are initialized with zero, and at each iteration new messages are computed in the following way,

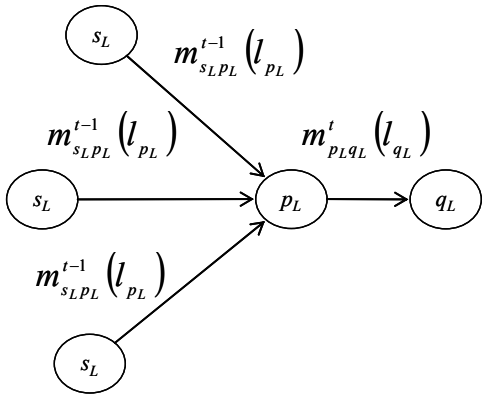


Figure 2.2: Message passing mechanism in Belief Propagation (BP); p_L sends message to q_L.

$$m^t_{q_L p_L}(l_{p_L}) = \min_{l_{p_L}} \left(D_{q_L}(l_{q_L}) + \lambda V(l_{q_L}, l_{p_L}) + \sum_{s_L \in \mathcal{N}(q_L) \setminus p_L} m^{t-1}_{s_L p_L}(l_{q_L}) \right), \tag{12}$$

where $\mathcal{N}(q_L) \setminus p_L$ denotes the neighbors of p_R excluding p_L. After T iterations, a belief vector is computed for each node as

$$b_{p_L}(l_{p_L}) = D_{p_L}(l_{p_L}) + \sum_{q_L \in \mathcal{N}(p_L)} m^T_{q_L p_L}(l_{p_L}). \tag{13}$$

Finally, the labels $l^*_{p_L}$ that minimize $b_{p_L}(l_{p_L})$ individually at each node are selected as the disparity map. To allow discontinuities, a function similar to (10) was used in the first BP algorithm for stereo proposed by Sun et al. [Sun03], where, rather than truncating the linear cost they had a function that changed smoothly from being almost linear near the origin to a constant value as the cost increases.

Several variants of BP exist [Wainwright05, Meltzer05, Kolmogorov06, Klaus06, Yang06c, Onofrio04]. In [Wainwright05, Meltzer05, Kolmogorov06], the authors tried to further develop or establish the theoretical aspects of belief propagation. In [Wainwright05], Wainright et al. proposed a variant of BP called tree reweighted belief propagation (TRBP) which differed slightly from BP in message update technique. TRBP is able to compute lower bound of a graph with cycles by decomposing the graph into trees. In [Meltzer05], Meltzer et al. experimentally showed that TRBP finds a global minimum lower than the graph cuts' and BP's. Although TRBP is inspired by the problem of finding the true global minimum, Kolmogorov showed that in practice, TRBP algorithms do not always provide a lower bound on the energy; also, it does not converge always [Kolmogorov06] either. He proposed a convergent tree reweighted message passing algorithm, which found the aggregated cost guaranteed not to increase. Experimental results demonstrated that on certain synthetic and real problems, this algorithm outperforms both the ordinary belief propagation and the tree-reweighted

algorithm in [Wainwright05]. In addition, on stereo problems with Potts interactions, he obtained the aggregated cost lower than the graph cuts' and BP's.

In [Klaus06, Yang06c], the performances of belief propagation were improved rather using image cues. In [Klaus06], the idea was to segment the left image based on color and assume each segmented part as a plane. The color segmentation improved matching around the discontinuous regions. With the same objective, in [Yang06c] the data cost was obtained from color weighted cross-correlation. [Onofrio04] is an application of belief propagation for modeling 3D face, where the data cost is obtained from area based estimate.

The standard belief propagation propagates the messages around a graph with loops. Therefore, the MAPs are redundant in the estimated cost, i.e., the disparity map is not obtained from an optimum on approximated cost. An experiment conducted by Szeliski et al. showed that, at best, BP gives an energy that is 3.4% higher and at worst 30% [Szeliski06]. The top ranked stereo matching algorithms listed in the performance table of Middlebury stereo site are BP and their variants [Middlebury].

2.1.4 Dynamic programming

Dynamic programming is an optimization method of solving problems that exhibit the properties of overlapping sub-problems and optimal substructures. Optimal substructure means that optimal solutions of sub-problems can be used to find the optimal solutions of the overall problem. For instance, in a graph with cost weighted edges, if cost of a path between two vertices U and V are optimal and W is any vertex on the path then the costs of UW and WV are also optimal. This concept can be applied to solve stereo matching problem by taking advantage of the sites for being on the same scan lines in rectified stereo images. If we match all the sites in row j in the left image with all the sites in row j in the right image we have a 2D array of matching costs. Costs of the corresponding pixels will stay approximately along the diagonal of this array. Stereo matching then becomes a problem of finding an optimal path on a two-dimensional search plane. Successful attempts at dynamic programming for solving the stereo correspondence problem are reported by Baker and Binford in 1981 [Baker81] and Ohta and Kanade in 1985 [Ohta85]. In the former, correlation was performed at edge points to reliably spot a few vertices on the path. Then, dynamic programming was applied to fill the remaining vertices on the path. In the later, additionally consistency between scan lines was maintained by enforcing continuity of vertical edges.

A two pass dynamic programming algorithm applying eq. (6) was proposed by Kim et al. [Kim05] to estimate cost of a site in a dynamic programming algorithm. In one of the passes, matching was performed within one scan line; in the other, consistency between scan lines was enforced. Data term was sum of absolute differences of intensities in homogeneous regions and sum of absolute differences of intensities convolved with Gaussian kernel in heterogeneous regions. In pass one, only left and right neighbors of p_L were included in the smoothness term. In pass two, neighbors from scan lines above and below the current scan line were included and aggregated with the cost estimated in pass one. For the smoothness term, a modified Pott model (Eq. (14)) allowed slanted surface

reconstruction to have the stair effect avoided. Discontinuity in disparity was maintained by setting the smoothness parameter λ inversely proportional to the intensity gradient.

$$V(l_{pL}, l_{qL}) = \begin{cases} 0 & \text{if } l_{pL} = l_{qL} \\ 0.5 & \text{if } |l_{pL} - l_{qL}| = 1. \\ 1 & \text{otherwise} \end{cases} \quad (14)$$

To gain speed in matching, some dynamic programming algorithms avoid energy function. A fast and automatic stereo matching algorithm based on dynamic programming was proposed by Benshair et al. [Benshair96]. Instead of using a cost function, threshold value on the data term was used to determine the optimal path. In a similar threshold based algorithm, Fortsmann et al. obtained real-time performance (30 FPS on a 2.2GHz PC) on dense matching combining coarse to fine approach and compiler optimization [Fortsmann04].

2.2 Middlebury College test bed for comparative studies

Middlebury College stereo test bed is a worldwide known website [Middlebury] where standard test stereo images with ground truth disparity maps and occlusions are posted. The images are captured with narrow baseline stereo rig and they have varieties structures in regards with color, depth variation, and shapes and sizes of surfaces. Each stereo pair has known ground truth disparity map obtained from laser scan. Occlusions in the disparity maps are detected from re-projections of the disparity maps and discontinuous regions in the disparity solutions are labeled. The test bed currently suggests a data set of four test stereo image pairs, also known as the new dataset, for evaluation: Tsukuba, Venus, Teddy, and Cones. Previously, they had a different data set of four images, also known as the old dataset: Tsukuba, Venus, Map, and Sawtooth. Apparently, these two datasets have the two image pairs, Tsukuba and Venus, in common.

Figure 2.3 shows one of these stereo pairs and its ground truth disparity and occlusion maps. In the ground truth images, the brightest point is the closest to the cameras. Figure 2.3(a) and (b) are the left and right stereo images. Notice that all points in the right image shift to the left (the visibility constraint). Disparities of the pixels are scaled by a number and then rounded to the nearest integer to visualize as shown in Figure 2.3(c). The scale factor is chosen to be such that after scaling, the maximum disparity does not pass 255. 10 pixels wide regions near the border of the image are shown black; they are not counted in the evaluation. Some portions in the left image are occluded in the right image. They are shown in black in Figure 2.3(e). As mentioned earlier, occlusions occur near the border of two surfaces. Figure 2.3(f) shows the discontinuous regions in white. Figure 2.4 shows all the test images used for evaluation by Middlebury College. Only one image of each image pair is shown in the figure.

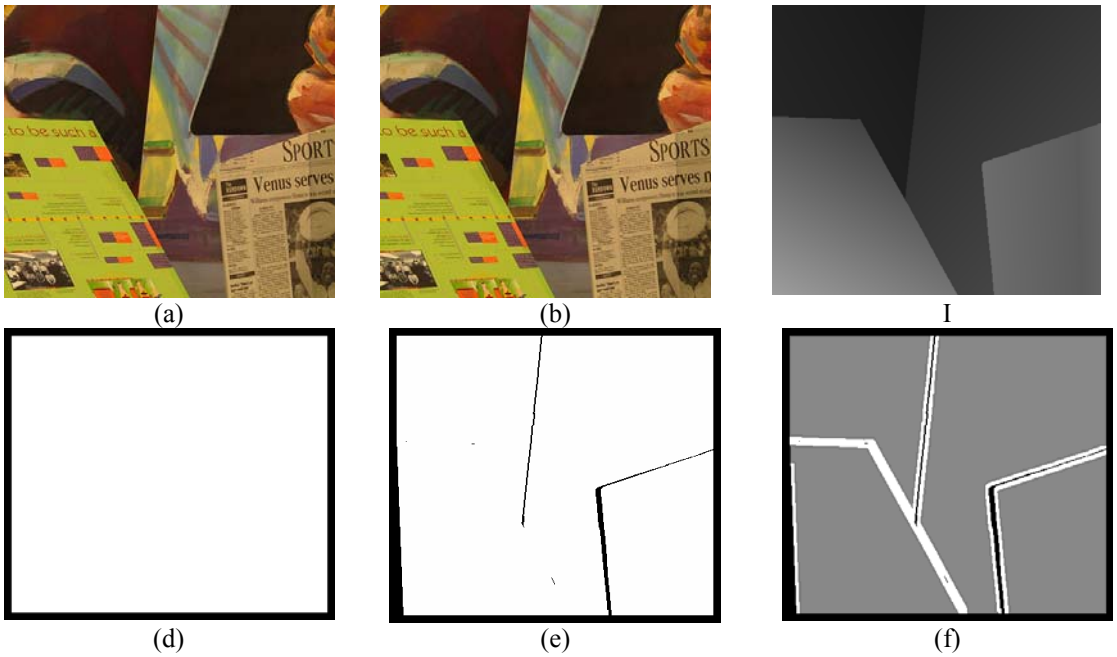
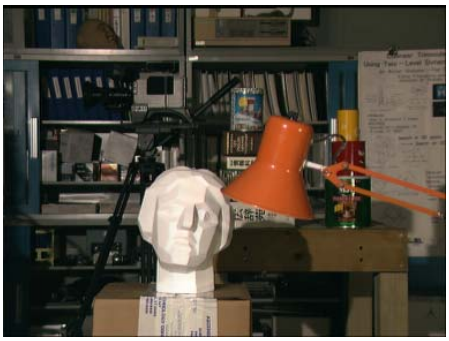


Figure 2.3: One of the Middlebury test stereo image pairs (image size 383×434), (a) left image, (b) right image, (c) ground truth disparity map, (d) 10 pixels of border excluded from performance evaluation, (e) regions of the left image that are occluded in the right image, and (f) discontinuous regions of the left image [Middlebury].



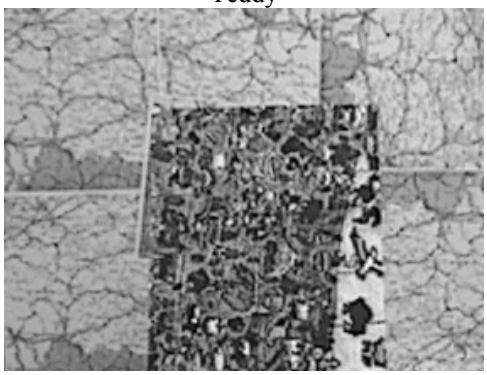
Tsukuba

Venus

Cone

Teddy

Sawtooth



Map

Figure 2.4: Ground truth test images used by Middlebury College for evaluation of stereo matching algorithms [Middlebury].

Middlebury College also posts a performance table online [Middlebury]. This table lists performance of many stereo matching algorithms published mostly after year 2000. Table 2.1 is prepared from the current on-line version of the evaluation table first published in the paper [Scharstein02] in 2002. Table 2.1 compares matching rates of competing MRF stereo matching algorithms against four test image pairs of the new dataset. For evaluation according to the rules set by the Middlebury College, all algorithms in Table 2.1 are run with constant parameter settings for all four images.

Each cell in the table has two numbers. The top number is the matching error obtained from comparison with the ground truth disparity map. These numbers represent the error percentages of bad pixels, i.e., pixels that have absolute disparity error greater than 1. For each image pair, the table reports error percentages for (1) non-occluded ('non-occ') pixels in the images, (2) all pixels ('all') including occlusions, and (3) pixels near depth discontinuities ('disc') (see Figure 2.3(f)). 10 pixels in Venus and 18 in Tsukuba near the border are ignored in the evaluation. For more details, we refer to the website, [Middlebury].

The bottom number in each cell is *rank*. The cumulative *score* of each algorithm in each column is indicated in the green column. The cumulative score is average of all the ranks of an algorithm. A rank is the position number of an algorithm in a list where all the algorithms are sorted in non-decreasing order according to a particular matching error rate. However, for a class-wise comparative study, the algorithms are listed according to their optimization classes in non-increasing order of overall performance. The algorithms mentioned in Table 2.2 are non-MRF algorithms. It is clear from these two tables that MRF algorithms show better performance than the non-MRF ones. Also, among MRF algorithms, BP based algorithms are the top performers.

2.3 Existing algorithms for MRF parameter estimations

Two major statistical techniques for automatic parameter estimation of the MRF parameters are published so far: one by Cheng and Caelli [Cheng07] and the other by Zhang and Seitz [Zhang07]. The former uses Markov Chain Monte Carlo (MCMC) technique for discontinuity handling, hence, sacrifices accuracy by not applying boundary values on the discontinuities. The later generates better matching results by applying the boundary values estimated with expectation maximization (EM) technique. We list performances of these two parameter estimation techniques in Table 2.3. In the table, the error statistics are estimated on two kinds of areas in the disparity solution: one is the non-occluded regions and the other is the discontinuous regions. Each cell in the table has two entries. The top entry indicates the percentage of matching error and a number below shows its rank in the Middlebury performance table for that matching error. A rank is the position of the algorithm when all algorithms are sorted in a non-decreasing order according to their error rates. All eight ranks are averaged to obtain the accumulated

score of an algorithm. The table shows that clearly, Zhang-Seitz algorithm performs better than the Cheng-Caelli algorithm.

Zhang-Seitz algorithm estimates the MRF parameters for data, smoothness, and discontinuity handling thresholds, iteratively and alternately with the matching iterations. In the algorithm, the matching algorithm starts with a set of initial values for the parameters and finds the disparity map. The parameters are updated from the current disparity map using EM iterations and then applied again to refine the disparity map. Since the EM iterations are nested inside the matching iteration, computational cost of the overall matching algorithm is high. Besides, although the MRF parameters are well defined and are seen to meet the convergence at optimality, disparity solution along the surface borders is not as good as expected.

2.4 Existing occlusion filling algorithms

Regions of one of the stereo images that are invisible in the other are called occlusions. Occlusions occur near the image borders. They also always appear inside the images when two or more distinct surfaces appear as foregrounds and backgrounds in the scene. In the literature of stereo vision, occlusion problem has been neither studied systematically nor solved properly, although, occlusion filling is important for many rendering applications. Many of the stereo matching algorithms treat occlusion as noise during the matching and therefore, do not detect or fill occlusions [Kim06]. In pixel-wise dense matching, such as in MRF stereo algorithms, detection of occlusions are performed implicitly [Klaus06, Sun05] or explicitly [Min08, Yang06c]. In implicit detection, occlusions are handled during the matching process. In the explicit ones, occlusions are detected performing the matching both ways and then comparing the two disparity values of corresponding points [Min08, Yang06c]. Klaus et al. segmented disparity planes of similar color intensities iteratively using mean shift algorithm and from analysis of matching costs. Disparity of an occluded point was extrapolated from disparity of the disparity plane. Since matching cost is used for segmentation of the plane, the occlusion filling algorithm is not independent of the matching algorithm and therefore, can not be applied to other algorithms. In addition, segmentation of the planes does not exclude occluded points from the segmentation process, i.e., occlusion filling is corrupted by initial inaccuracies in disparities of the occluded points. Sun et al. detected occlusion during the matching process and filled them with disparities of non-occluded neighbors [Sun05]. Thus, slope of the plane was not considered in estimation of disparity of the occluded point. Min and Sohn [Min08] detected occlusions explicitly from both way matching. The occlusion was estimated from diffusion of energies of neighboring non-occluded points. The energies were obtained from the matching algorithm and therefore, occlusion filling was not independent of the matching algorithm. Also, slope of the surface the occluded point belonged to was not considered in the estimation. Yang et al. filled disparities of occluded points by directly assigning them the disparities of non-

occluded neighbors [Yang06c]. Table 2.4 summarizes merits and drawbacks of occlusion filling methods used in some of the existing algorithms.

Table 2.1: Performance table of MRF stereo matching algorithms

Class	Algorithm	Score	Tsukuba			Venus			Teddy			Cones			Parameter tuning? (Manual, Heuristic, Statistical)
			non-occ	all	disc	non-occ	all	disc	non-occ	all	disc	non-occ	all	disc	
Hybrid (BP)	AdaptingBP [Klaus06]	2.1	1.11 6	1.37 3	5.79 7	0.10 1	0.21 2	1.44 1	4.22 4	7.06 2	11.8 4	2.48 1	7.92 2	7.32 1	Manual
Hybrid (BP)	DoubleBP [Yang06c]	4.8	0.88 2	1.29 1	4.76 1	0.14 5	0.60 13	2.00 7	3.55 3	8.71 5	9.70 2	2.90 4	9.24 10	7.80 3	Manual
Hybrid (BP)	SubPixDoubleBP [Yang07]	5.5	1.24 10	1.76 13	5.98 8	0.12 2	0.46 6	1.74 4	3.45 1	8.38 4	10.0 3	2.93 5	8.73 6	7.91 4	Manual
Hybrid (BP)	SymBP+occ [Sun05]	10.6	0.97 4	1.75 12	5.09 3	0.16 6	0.33 3	2.19 8	6.47 8	10.7 6	17.0 14	4.93 23	10.7 15	10.9 14	Manual
Hybrid (BP)	OverSegmBP [Zitnick07]	14.0	1.69 22	1.97 17	8.47 23	0.51 17	0.68 15	4.69 18	6.74 10	11.9 11	15.8 8	3.19 8	8.81 8	8.89 11	Manual
Hybrid	C-SemiGlob [Hirschmüller06]	12.1	2.61 28	3.29 23	9.89 26	0.25 12	0.57 10	3.24 15	5.14 6	11.8 8	13.0 6	2.77 2	8.35 4	8.20 5	Manual + Heuristic
Hybrid (BP)	EnhancedBP [Larsen07]	16.0	0.94 3	1.74 10	5.05 3	0.35 15	0.86 17	4.34 17	8.11 21	13.3 18	18.5 21	5.09 25	11.1 22	11.0 20	Manual
Hybrid	SemiGlob [Hirschmüller07]	18.8	3.26 31	3.96 28	12.8 33	1.00 22	1.57 21	11.3 27	6.02 2	12.2 14	16.3 10	3.06 7	9.75 13	8.90 12	Manual + Heuristic
Hybrid (BP)	RealtimeBP [Yang06b]	21.9	1.49 19	3.40 25	7.87 21	0.77 19	1.90 25	9.00 26	8.72 25	13.2 17	17.2 15	4.61 21	11.6 18	12.4 21	Manual
Hybrid (BP)	Layered [Zitnick04]	23.7	1.57 20	1.87 16	8.28 22	1.34 25	1.85 23	6.85 22	8.64 24	14.3 22	18.5 22	6.59 30	14.7 29	14.4 29	Manual
Hybrid (BP)	RealTimeGPU [Wang06]	21.3	2.05 25	4.22 30	10.6 29	1.92 31	2.98 29	20.3 34	7.23 15	14.4 23	17.6 17	6.41 29	13.7 28	16.5 31	Manual
Hybrid (GC)	GC+occ [Kolmogorov01]	23.2	1.19 7	2.01 19	6.24 10	1.64 28	2.19 27	6.75 20	11.2 30	17.4 30	19.8 27	5.36 26	12.4 27	13.0 28	Manual
Hybrid (GC)	MultiCamGC [Kolmogorov02]	24.0	1.27 13	1.99 18	6.48 12	2.79 33	3.13 31	3.60 16	12.0 31	17.6 31	22.0 30	4.89 24	11.8 25	12.1 24	Manual
Hybrid (GC)	GC [Boykov01]	30.2	1.94 23	4.12 29	9.39 25	1.79 30	3.44 32	8.75 25	16.5 36	25.0 37	24.9 33	7.70 31	18.2 32	15.3 30	Heuristic

Class	Algorithm	Score	Tsukuba			Venus			Teddy			Cones			Parameter tuning? (Manual, Heuristic, Statistical)
			non-occ	all	disc	non-occ	all	disc	non-occ	all	disc	non-occ	all	disc	
Local (ICM)	Segm+visib [Bleyer04]	11.8	1.30 15	1.57 5	6.92 18	0.79 20	1.06 18	6.76 21	5.00 5	6.54 1	12.3 5	3.72 12	8.62 5	10.2 16	Manual
Local	AdaptWeight [Yoon06]	17.0	1.38 17	1.85 15	6.90 17	0.71 18	1.19 19	6.13 19	7.88 19	13.3 19	18.6 23	3.97 18	9.79 14	8.26 6	Manual
Local	CostRelax [Brockers05]	27.6	4.76 35	6.08 34	20.3 37	1.41 27	2.48 28	18.5 33	8.18 22	15.9 27	23.8 31	3.91 16	10.2 18	11.8 23	Manual
DP	RegionTreeDP [Lei06]	15.1	1.39 18	1.64 8	6.85 16	0.22 8	0.57 10	1.93 6	7.42 16	11.9 10	16.8 13	6.31 28	11.9 26	11.8 22	Manual
DP	ReliabilityDP [Gong05]	28.5	1.36 16	3.39 24	7.25 19	2.35 32	3.48 33	12.2 30	9.82 28	16.9 28	19.5 26	12.9 37	19.9 31	19.7 28	Manual
DP	TreeDP [Veksler05]	29.4	1.99 24	2.84 22	9.96 27	1.41 26	2.10 26	7.74 24	15.9 35	23.9 35	27.1 36	10.0 33	18.3 33	18.9 32	Manual
DP	SegTreeDP [Deng06]	16.9	2.21 26	2.76 21	10.3 28	0.46 16	0.60 12	2.44 10	9.58 26	15.2 25	18.4 20	3.23 9	7.86 1	8.83 9	n/a
DP	DP [Intlille94]	33.8	4.12 33	5.04 33	12.0 31	10.1 40	11.0 40	21.0 35	14.0 32	21.6 32	20.6 28	10.5 34	19.1 29	21.1 29	n/a
DP	SO [Christopher06]	37.3	5.08 37	7.22 38	12.2 32	9.44 39	10.9 39	21.9 36	19.9 39	28.2 40	26.3 35	13.0 38	22.8 39	22.3 36	Manual

Table 2.2: Performance table of non-MRF stereo matching algorithms

Algorithms	Score	Tsukuba			Venus			Teddy			Cones		
		non-occ	all	disc	non-occ	all	disc	non-occ	all	disc	non-occ	all	disc
SO+borders [Mattoccia07]	12.4	1.29 14	1.71 9	6.83 15	0.25 13	0.53 9	2.26 9	7.02 12	12.2 9	16.3 9	3.90 14	9.85 15	10.2 17
DistinctSM [Yoon07]	13.7	1.21 9	1.75 11	6.39 11	0.35 14	0.69 16	2.63 13	7.45 17	13.0 16	18.1 18	3.91 15	9.91 17	8.32 7
SegmentSupport [Tombari07]	14.6	1.25 11	1.62 7	6.68 13	0.25 11	0.64 14	2.59 12	8.43 23	14.2 21	18.2 19	3.77 13	9.87 16	9.77 15
TensorVoting [Mordohai06]	26.7	3.79 32	4.79 32	8.86 24	1.23 24	1.88 24	11.5 28	9.76 27	17.0 29	24.0 32	4.38 20	11.4 23	12.2 25
PhaseBased [Etriby07]	35.2	4.26 34	6.53 35	15.4 35	6.71 36	8.16 36	26.4 38	14.5 33	23.1 33	25.5 34	10.8 36	20.5 37	21.2 35
SSD+MF [Scharstein02]	35.6	5.23 38	7.07 36	24.1 38	3.74 34	5.16 34	11.9 29	16.5 37	24.8 36	32.9 38	10.6 35	19.8 35	26.3 37
STICA [Audirac05]	36.8	7.70 39	9.63 40	27.8 39	8.19 37	9.58 37	40.3 40	15.8 34	23.2 34	37.7 39	9.80 32	17.8 31	28.7 39
PhaseDiff [Etriby06]	38.0	4.89 36	7.11 37	16.3 36	8.34 38	9.76 38	26.0 37	20.0 40	28.0 39	29.0 37	19.8 40	28.5 40	27.5 38
Infection [Olague06]	38.4	7.95 40	9.54 39	28.9 40	4.41 35	5.53 35	31.7 39	17.7 38	25.1 38	44.4 40	14.3 39	21.3 38	38.0 40

Table 2.3: Percentages of matching error of existing parameter estimation algorithms for the Middlebury test images – Tsukuba, Venus, Map, and Sawtooth; the matching algorithm used is BP

Techniques	Error statistics regions	Tsukuba	Venus	Map	Sawtooth	Score
Zhang-Seitz [Zhang07]	Non-occluded	1.87 20	1.53 20	0.20 2	0.83 14	12.0
	Discontinuous	7.13 9	10.37 21	2.20 1	3.48 9	
Cheng-Caelli [Cheng07]	Non-occluded	3.65 28	3.41 32	0.10 1	1.23 21	21.1
	Discontinuous	15.33 32	17.40 29	1.33 1	7.91 25	

Table 2.4: Occlusion filling strategies in the a few state of the art stereo matching algorithms

Papers	Occlusion filling strategies	Working principle and drawbacks
[Min08]	Diffusion in Intensity	[Min08] assigns disparity to the occluded point from non-occluded points with similar color intensities. Occlusion filling does not consider slope of the planes belonging to the occluded point.
[Klaus06]	Implicit filling during the matching process	Occlusion filling is done implicitly during the matching process by segmenting planes with similar color intensities. Occlusion filling of occluded planes is undefined. The occlusion filling algorithm is not independent to matching.
[Yang06]	Neighbor's disparity assignment	Occlusion filling does not account for the slope of the plane that belongs to the occluded point.
[Sun05]	Neighbor's disparity assignment	Occlusion filling does not account for the slope of the plane that belongs to the occluded point.
[Kim05]	Occlusion is treated as noise	Occlusion is neither detected nor filled but treated as noise during the matching process.

2.5 Stand of our stereo matching algorithm in stereo vision

In a top-to-bottom structure, Figure 2.5 describes the relative position of the proposed stereo matching algorithm in the literature of stereo matching in computer vision. All stereo matching algorithms can be classified into sparse and dense matching. Sparse matching is mostly area based and many of them are developed in early period of stereo vision. Our matching algorithm is dense. Dense matching algorithms can be divided into two categories: energy minimizing and area based. We have already reviewed the energy minimizing classes, where many of the algorithms are also developed in MRF. Many dynamic programming algorithms use spatial information applying neighborhood in matching, this is why dynamic programming is included in energy minimizing class.

Figure 2.5 describes that belief propagation, a hybrid algorithm, was first proposed by Sun et al. in [Sun03], where parameters were determined manually. Later in 2007, statistical estimation algorithm for the parameters was proposed by Zhang and Seitz in [Zhang07] and Cheng and Caelli in [Cheng07]. All the other algorithms in energy minimizing class use heuristic or manual determination of the parameters. In contrast, in the proposed stereo matching algorithm in this dissertation, the parameters are estimated automatically from statistics. As mentioned earlier, the proposed parameter estimation technique also works with BP offering some efficiency in computation.

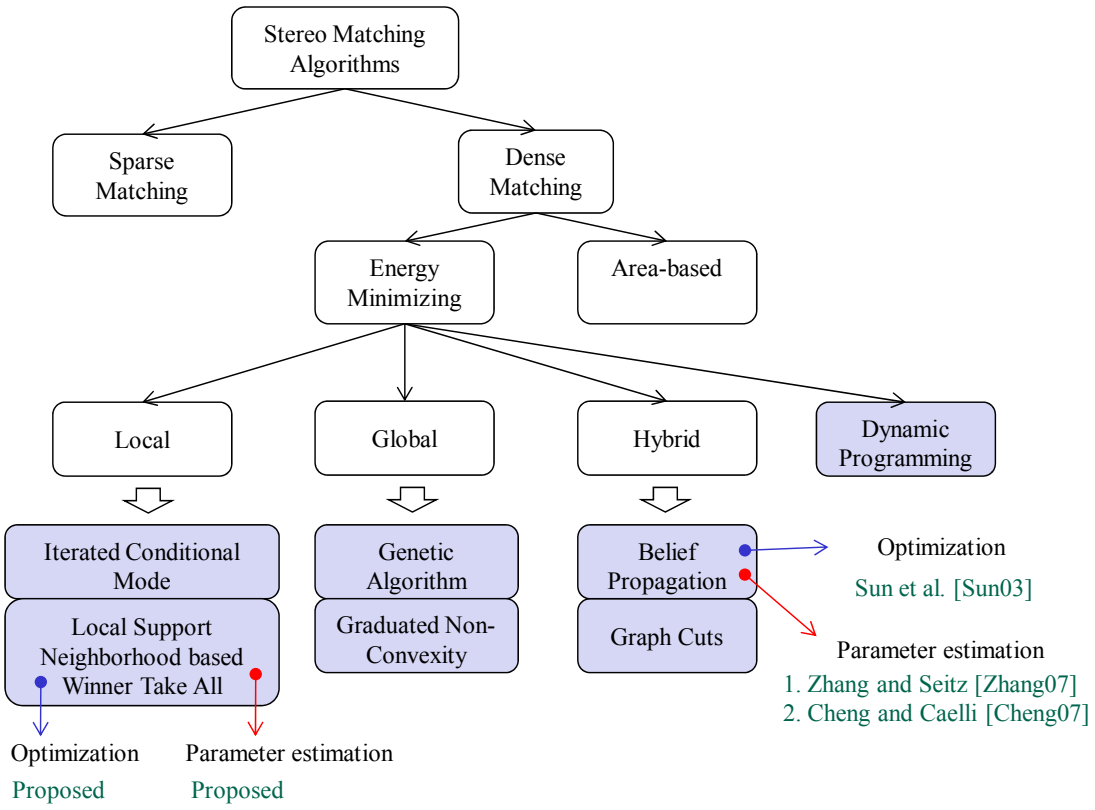


Figure 2.5: Demonstration of the stand of the proposed stereo matching work of this dissertation in a top to bottom breakdown chart of the stereo matching algorithms

3 MRF stereo matching

In this chapter, we describe the proposed MRF stereo matching algorithm that performs dense matching. In dense matching, all pixels in the images are matched. The advantages with pixel based matching are that estimating the data cost only from one pixel reduces the computational time significantly and at the same time discontinuity and occlusion detections can be handled appropriately. The price one has to pay is an increased locality of the solution. Depending on an initial labeling, the solution can settle to a much worse final labeling in an increased number of local minima. However, the number of local minima can be reduced by increasing the number of neighbors in $\mathcal{N}(p_L)$ in eq. (6) or (7). For a global optimization of the aggregated cost function we incorporate an adaptive support neighborhood based matching. In the cost function, we handle discontinuity in smoothness by introducing some parameters. The adaptability of the support neighborhood system provides additional consistency in handling discontinuity in smoothness.

There are four kinds of basic parameters in our algorithm, all estimated statistically. Two kinds of the parameters are for the data model and the others are for the smoothness. One kind of the data model parameters is for modeling data likelihood, namely ΔI_{LR}, which is the absolute intensity difference between corresponding points in both images and the other kind is for occlusion (i.e. discontinuity in data likelihood) handling. Two kinds of parameters are defined for the smoothness model. One kind is for modeling smoothness, namely Δl_{LL}, which is the absolute difference between the disparities of two neighboring pixels in the left image and the other is for handling discontinuity in disparity.

In the following sections, first we model the cost function, which is done by relaxing the neighborhood, $\mathcal{N}(p_L)$. Then, we introduce the data and smoothness models with their parameters. These models are developed from evidences in the ground truth datasets. Next, we introduce estimation techniques of all the parameters. Then, an adaptive support neighborhood based mechanism for optimization is described. Finally, we propose a pair of symmetric cost functions for improved matching and occlusion detection.

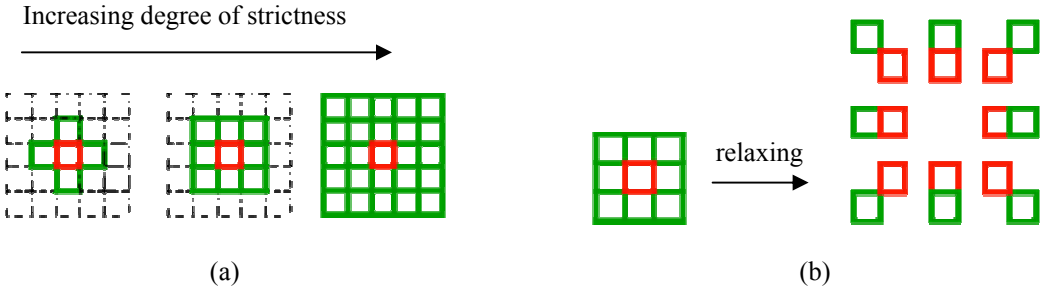


Figure 3.1: (a) Demonstration of strictness of a neighborhood and (b) relaxing the neighborhood strictness by pairing data with each neighbor.

3.1 Cost function by relaxing the neighborhood

Definition of neighborhood, $\mathcal{N}(p_L)$, in MRF stereo matching literature is confined with modeling assumption. The smallest neighborhood is a first order 4-neighbor system; the largest could be fairly large depending on the assumption. An important concept we introduce in this section is *strictness* of a neighborhood – a large neighborhood is stricter than a small neighborhood. Thus, increasing the number of neighbors imposes increasing *strictness* (see Figure 3.1). What we intend to do is to relax this strictness. Relaxation is performed by pairing the data point with each neighbor.

Relaxing a neighborhood has some important consequences on the cost function. The summation sign comes in the front and operates on both the data and smoothness terms making them work into the same dimension of probability (eq. (15)).

$$E_{p_L}(l_{p_L}) = \sum_{q_L \in \mathcal{N}(p_L)} \lambda_D D_{p_L}(l_{p_L}) + \lambda_{q_L} V(l_{p_L}, l_{q_L}). \tag{15}$$

For direct estimation of the parameters, we use eq. (5) with λ_D as the data model parameter and λ_{q_L} as the smoothness model parameter. Notice that taking the parameters λ_D and λ_{q_L} into account balancing between the data and smoothness costs in eq. (15) occurs giving both the costs equal importance, namely, 50% of weight to each. This is how our modeling of the cost function in eq. (15) differs from the widely used eq. (6), where the cost function is parameterized with one combined parameter, λ.

3.2 Modeling the data likelihood and smoothness

In this section we model the data likelihood. First, we show that the data likelihood can be modeled with exponential distribution. Second, we show that using Zhang-Seitz estimation technique [Zhang07], the data likelihood parameters can be estimated *a priori*

from one image applying a noise equivalence hypothesis that will be described into details shortly. Note that the estimations of the data and smoothness parameters are interdependent – one of them can not be estimated unless the other one is estimated. *A priori* estimation of the data likelihood parameters eliminates this interdependency and thus, convergence of the estimation of the parameters, hence of the matching algorithm, is ensured.

In modeling the data likelihood, $\Delta I_{LR} = |I(p_L) - I(p_R)|$ is the random variable we want to find distribution of; here, $I(p_L)$ is the intensity map of p_L and p_R is a point in the right image corresponding to p_L. To find the closest distribution ΔI_{LR} can be described with, we build frequency histograms from ground truth data provided online by the Middlebury College. The frequency diagrams plot the numbers of counts of each ΔI_{LR} value for a particular stereo image pair; apparently, the condition $0 \le \Delta I_{LR} \le 255$ holds. Figure 3.2 shows the frequency diagrams and their corresponding theoretical models for two distributions – exponential and Gaussian. In the study of exponential, $\Delta I_{LR} = |I(p_L) - I(p_R)|$ is the random variable and in Gaussian, we consider $\Delta I_{LR} = I(p_L) - I(p_R)$ as the random variable.

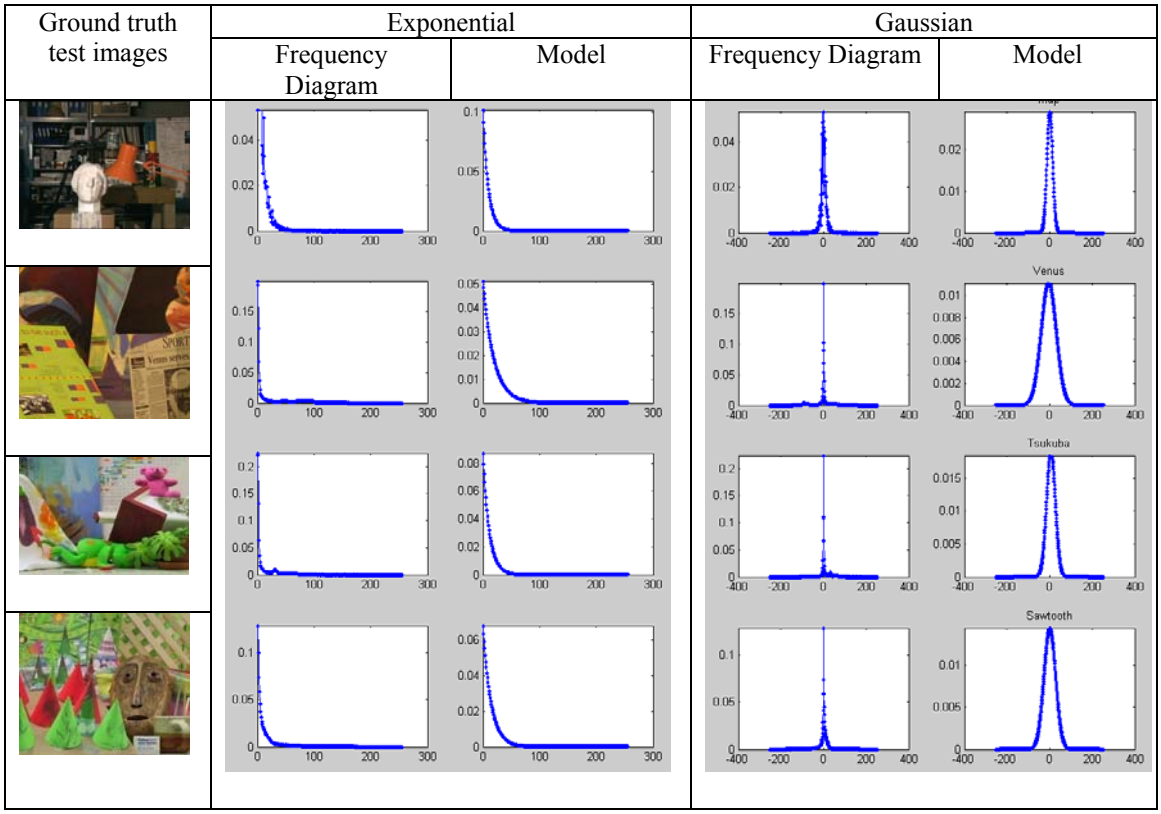


Figure 3.2: Frequency distribution and model of ΔI_{LR} obtained from Middlebury ground truth stereo images.

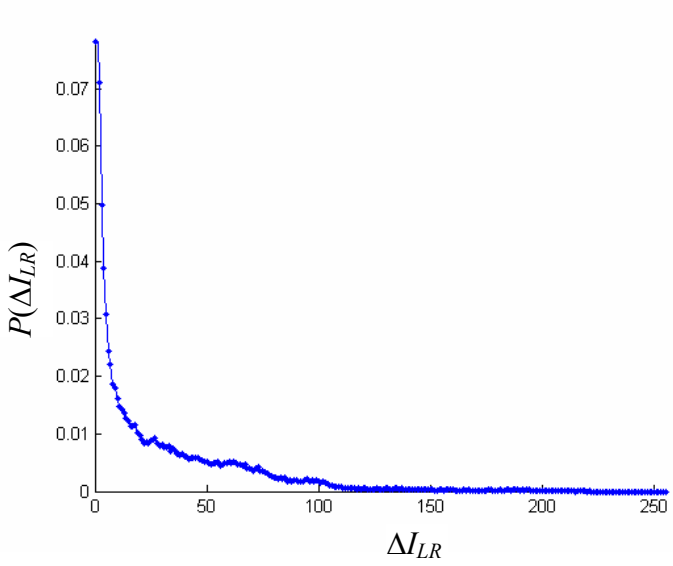


Figure 3.3: Zoomed in frequency distribution of ΔI_{LR} obtained from Middlebury ground truth stereo image 'Tsukuba'.

Visually, the plots suggest that an exponential distribution fits better than a Gaussian. If we look into one of the frequency histograms closely, this visual reasoning for exponential distribution is justified. Figure 3.3 shows a zoomed-in frequency histogram (normalized) built from the image pair Tsukuba. Clearly, we observe two important shape features in this plot: sharp pick and fat tail. Both of these features are described well by exponential distributions. The fat tail is due to violation of the brightness constancy apparently caused by a variation in the angle a scene point makes with the image planes while capturing images from two viewpoints.

3.2.1 Data likelihood in Zhang-Seitz algorithm

For the variable $\Delta I_{LR} = |I(p_L) - I(p_R)|$, where p_R is the right image point corresponding to p_L, the following mixture model with exponential and uniform distributions is defined:

$$P(\Delta I_{LR}) = \begin{cases} \zeta_{LR} \exp(-\mu_{LR} \Delta I_{LR}), & \text{if } p_L \text{ has a match} \\ \dfrac{1}{N_{LR}}, \text{otherwise, i.e., } p_L \text{ is occluded} \end{cases},$$

where μ_{LR} is a decay rate, ζ_{LR} is the normalizing factor with $\zeta_{LR} = \dfrac{(1 - \exp(-\mu_{LR}))}{(1 - \exp(-\mu_{LR} N_{LR}))}$, and $N_{LR} = \max(\Delta I_{LR}) + 1$. Say that probability of ΔI_{LR} having the exponential distribution is α_{LR}. Then, in mixture model,

$$P(\Delta I_{LR}) = \alpha_{LR}\zeta_{LR}\exp(\mu_{LR}\Delta I_{LR}) + (1-\alpha_{LR})\frac{1}{N_{LR}}. \quad (16)$$

The model needs to be mixture because, in $\Delta I_{LR} = |I(p_L) - I(p_R)|$, if p_L is an occluded point, i.e. there is no p_R that corresponds to p_L, then ΔI_{LR} follows a uniform distribution; otherwise it follows the exponential distribution as described in the beginning of Section 3.2. For estimation of the MRF parameter, μ_{LR}, an EM algorithm was adopted in [Zhang07].

3.2.2 Data likelihood parameters with Noise Equivalence

Hypothetically, we find that ΔI_{LR} can be modeled from one of the images assuming that both the stereo images are taken with identical cameras. Say q_L is a neighbor of p_L in the left image. q_L is taken in any direction but the direction remains the same for any p_L. Our hypothesis is that the two random variables, $|I(p_L) - I(p_R)|$, i.e. the *between image* noise and $|I(p_L) - I(q_L)|$, i.e. the *within image* noise, follow the same distribution. This hypothesis enables us to have an estimate of the data model parameters *a priori* from one of the images and eliminate a nested iteration of the alternating optimization algorithm proposed in [Zhang07, Cheng07]. We estimate the parameters iteratively only once before the matching iteration starts.

We apply the mixture model described in section 3.2.1 to $\Delta I_{LL} = |I(p_L) - I(q_L)|$, which consequently gives us estimates of the data likelihood parameters according to the noise equivalence hypothesis proposed above. Thus, the new mixture model for an estimate of μ_{LR} can be defined as the following:

$$P(\Delta I_{LL}) = \begin{cases} \zeta_{LL}\exp(-\mu_{LL}\Delta I_{LL}), & \text{if } p_L \text{ and } q_L \text{ are in homogeneous region, i.e., gradient is due only to noise.} \\ \dfrac{1}{N_{LL}}, & \text{otherwise, i.e., } p_L \text{ and } q_L \text{ have gradient due to texture difference.} \end{cases}$$

Here, μ_{LL} is the decay rate equivalent to μ_{LR} and ζ_{LL} is the normalizing factor with $\zeta_{LL} = \dfrac{(1-\exp(-\mu_{LL}))}{(1-\exp(-\mu_{LL}N_{LL}))}$, and $N_{LL} = \max(\Delta I_{LL}) + 1$. Except for the subscripts, we keep the notations the same for clarity. Say, probability that p_L has a homogeneous neighbor q_L, taken always in one particular direction, is β_{LL}. Then, in mixture model,

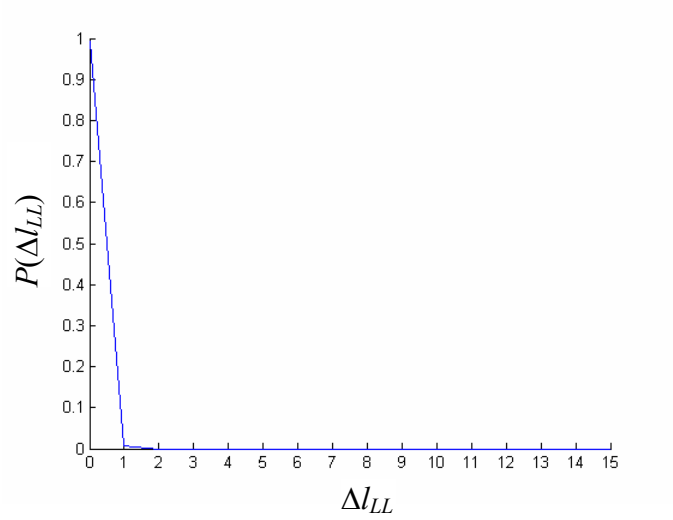


Figure 3.4: Frequency diagram of $\Delta l_{LL} = \left| l_{p_L} - l_{q_L} \right|$, where q_L is the immediate right neighbor of p_L (obtained from Middlebury ground truth stereo image, Tsukuba).

$$P(\Delta I_{LL}) = \beta_{LL} \xi_{LL} \exp(-\mu_{LL} \Delta I_{LL}) + (1 - \beta_{LL}) \frac{1}{N_{LL}}. \tag{17}$$

For the data term we take the variance of ΔI_{LL} as μ_{LL}^{-2}. Estimation of μ_{LL} is described in section 3.3.

3.2.3 Modeling smoothness

The prior we use is smoothness in labeling. The variable for smoothness between p_L and one of its neighbors q_L is defined with absolute difference of their labels, $\Delta l_{LL} = \left| l_{p_L} - l_{q_L} \right|$. We model Δl_{LL} with exponential distribution (eq. (18)) as

$$P(\Delta l_{LL}) = \nu_{q_L} \exp(-\nu_{q_L} \Delta l_{LL}), \tag{18}$$

where, ν_{q_L} is the decay rate that equals to λ_{q_L}.

Figure 3.4 shows a normalized histogram of Δl_{LL} when q_L is the immediate right un-occluded neighbor of p_L. The histogram is built from ground truth data for the stereo pair 'Tsukuba'. The histogram has sharp pick and its tail decays fast, turning to zero past $\Delta l_{LL} = 1.0$. Shape of the histogram of Δl_{LL} depends on the scene structures; generally, however, smoothness term in stereo (both in manual and statistical parameter setting

cases) is designed in absolute difference between neighboring disparities [Sun03, Boykov01]; thus, Δl_{LL} is consistent with exponential distribution.

3.3 Estimation of the parameters

The data likelihood model parameters are estimated from one of the stereo images using expectation maximization before the matching iteration starts. Then, an alternating optimizing algorithm estimates the smoothness model parameters using a combination of maximum likelihood and disparity gradient constraint. The prior model parameters are first estimated from the current map of disparity. Then, estimated values are applied to matching to obtain an improved disparity map. These two steps alternate until convergence in matching is achieved.

3.3.1 Data likelihood model parameters

Core derivations presented in this section for the estimation of data likelihood parameters using expectation maximization (EM) algorithm are due to [Zhang07] except that instead of using both the stereo images now we use any one of them. Expectation maximization (EM) is better suited for estimation of the parameters of a mixture models. First, we define an estimate of the probability of a particular site p_L becoming homogeneous with its immediate right neighbor q_L, whom we call ω_{p_L}, and we define b_L as a set of pixels in the rightmost column of S_L. ω_{p_L} is defined as the following,

$$\omega_{p_L} \overset{def}{=} \frac{\beta_{LL}\zeta_{LL}\exp\left(-\mu_{LL}\Delta I_{LL}\right)}{\beta_{LL}\zeta_{LL}\exp\left(-\mu_{LL}\Delta I_{LL}\right)+\dfrac{1-\beta_{LL}}{N_{LL}}}. \tag{19}$$

Note that q_L could be a neighbor in any direction, but the direction has to be always the same for all p_L. For EM operation, we take expected log-probability, $E_{\omega}\left(\log P\left(\Delta I_{LL}\middle|\beta_{LL}\right)\right)$, of (17), which is

$$\sum_{p_L\in\Omega}\omega_{p_L}\log\left(\beta_{LL}\xi_{LL}\exp\left(-\mu_{LL}\Delta I_{LL}\right)\right)+\left(1-\omega_{p_L}\right)\log\frac{1-\beta_{LL}}{N_{LL}}. \tag{20}$$

Eq. (20) can be differentiated with respect to β_{LL} and μ_{LL} for estimation of β_{LL} and μ_{LL}. EM assumes that the best event (i.e. occurrence of the best set of sample) has not happened yet and estimation of the parameters has to be iterative with each next iteration working with a better set of observations for more accurate estimation of the parameters.

According to Dempster et al. [Dempster77] and Jeff Wu [Wu83] the expected value in EM algorithm is non-decreasing in every iteration. However, the non-decreasing property does not guarantee convergence of the EM parameters unless certain conditions are met. In our case, Eq. (20) has two independent parameters, β_{LL} and μ_{LL}. Say, for the parameter vector $\phi = [\beta_{LL} \;\; \mu_{LL}]'$, $\phi^{(i)}$ indicates the values of β_{LL} and μ_{LL} at an iteration *i*. If $E_{\omega}^{(i)}$ is the value of $E_{\omega}\left(\log P\left(\Delta I_{LL} \middle| \beta_{LL}\right)\right)$ at iteration *i*, then ϕ converges if two conditions satisfy [Dempster77, Wu83]. The conditions are the following.

(1) the sequence $E_{\omega}^{(i)}$ is bounded and

(2) $E_{\omega}^{(i+1)} - E_{\omega}^{(i)} \geq \delta \left\| \phi^{(i+1)} - \phi^{(i)} \right\|$, for any $\delta > 0$ and all *i*.

In our case, the first condition holds, since our data variable, ΔI_{LL}, is bounded; however, it is not obvious if the second condition also holds. Experimentally, we have seen that the parameters β_{LL} and μ_{LL} converge reliably (see Figure 3.5). For more details on convergence of EM algorithms we refer to a text by Gelman et al. [Gelman03].

Differentiating (20) with respect to β_{LL}, we obtain the following estimate,

$$\beta_{LL} = \frac{1}{\left| S_L \setminus b_L \right|} \sum_{p_L \in S \setminus b_L} \omega_{p_L} , \tag{21}$$

and with respect to μ_{LL}, we obtain an estimate of μ_{LL} as the solution to the equation,

$$\frac{1}{1 - \exp(-\mu_{LL})} - \frac{1}{1 - \exp(-N_{LL}\mu_{LL})} = -\frac{\sum_{p_L \in S_L \setminus b_L} \omega_{p_L} \Delta I_{LL}}{\sum_{p_L \in S_L \setminus b_L} \omega_{p_L}}, \tag{22}$$

which has a close form,

$$\mu_{LL} = \log\left(\frac{\sum_{p_L \in S_L \setminus b_L} \omega_{p_L}}{\sum_{p_L \in S_L \setminus b_L} \omega_{p_L} \Delta I_{LL}} + 1 \right), \tag{23}$$

when *N* is big enough (which usually is) to ignore the second term in the left hand side of (22). The three data likelihood parameters for the R, G, and B color channels for Venus are estimated and plot in Figure 3.5. The plots show that the estimation algorithm converges.

Table 3.1 shows the values of μ_{LL}^{-1} estimated from one of the stereo images by applying the equivalence hypothesis and μ_{LR}^{-1} estimated from both images by applying standard BP with parameter estimation of [Zhang07] to establish the correspondence between p_L and p_R. The table shows that the parameter values are reasonably close. The

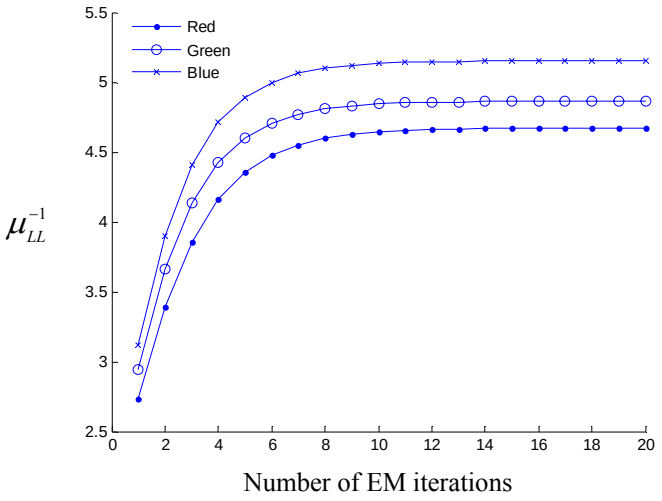


Figure 3.5: Image noise in three channels estimated from the left image 'Venus'

values of the both set of parameters mostly maintain the same color order when sorted and thus, provide additional prior evidence on the validity of the equivalence hypothesis. Now, we have the likelihood model parameter μ_{LL}^{-1} estimated before the matching iteration starts and thus, avoid the nested EM iteration.

3.3.2 Smoothness parameters

In estimation of the smoothness parameters we apply the disparity gradient constraint, $\Delta l_{LL}/\Delta x_{LL} \le T_{\Delta l_{LL}}$, for any $\Delta x_{LL} \ge 1.0$, where, Δx_{LL} is the absolute difference between the x-coordinates of p_L and q_L. We assign 1.0 to $T_{\Delta l_{LL}}$. The above definition of the disparity gradient, i.e. $\Delta l_{LL}/\Delta x_{LL}$, was established by Tyler in his work of psychophysical study for stereoscopic fusion in 1973 [Tyler73]. Tyler showed that fusion depended both on Δl_{LL} and Δx_{LL}. Later, another study by Burt and Julesz concluded that human stereo matching was constrained to disparity gradients ≤ 1.0 [Burt80]. Pollard

Table 3.1: μ_{LL}^{-1} estimated with and without applying noise equivalence

Variables	Color Channels	Tsukuba	Venus	Teddy	Cones
$\lvert I(p_L) - I(p_R)\rvert$	R	3.09	4.73	8.18	12.52
	G	3.18	3.68	6.42	11.61
	B	3.46	5.58	9.60	11.91
$\lvert I(p_L) - I(q_L)\rvert$	R	4.73	5.67	6.48	9.29
	G	4.92	4.77	5.85	8.69
	B	5.21	5.96	7.86	9.32

et al. further supported the study by stating that for most naturally occurring scene surfaces, including quite jagged ones, the disparity gradients between correct matches of image primitives lie within 1.0 [Pollard85]. More recently, the constraint $\Delta l_{LL}/\Delta x_{LL} \leq 1.0$ was successfully applied by Chen and Medioni in matching stereo images of human faces [Chen01] and by Oriot and Besnerais in matching aerial stereo images [Oriot98].

The estimation of the smoothness model parameter is performed by maximum likelihood (ML), hence single iteration, constrained with a composite condition formed applying both the disparity gradient constraint and 3*sigma confidence. The composite condition is defined as

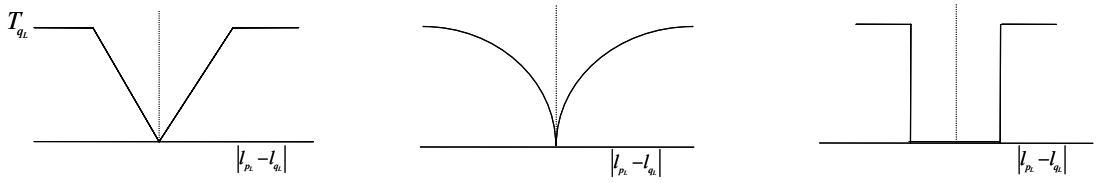
$$\Delta l_{LL} \leq \max\left(T_{\Delta l_{LL}},\, 3\nu_{q_L}^{-1}\right), \tag{24}$$

where Δx_{LL} is omitted, since the disparity gradient constraint is independent of Δx_{LL}. In ML estimation, any Δl_{LL} that does not satisfy the condition in (24) is considered outlier resulting from surface discontinuities. In our case, the ML estimation shows strong consistency, since, the exponential distribution in eq. (18) is regular up to its third derivative, i.e., the first, second, and third derivatives of $P(\Delta l_{LL})$ with respect to ν_{q_L} exist and our observations are I.I.D [Serfling80, Serfling01]. Strong consistency says that with the number of observations increasing to infinity, the parameter to be estimated approaches to the 'true' value. We have one observation for each pixel and large number of pixels (except the pixels at the top and right boundaries) in our estimation. Therefore, the estimation is reliable.

We estimate two *basic* smoothness model parameters for two neighbors of p_L. With respect to p_L, one of these neighbors is located at coordinate (1, 0); we call this neighbor q_{L10}. The other one is located at (0, 1), which we call q_{L01}. Let us say that the two basic parameters (standard deviations) related to those two neighbors are ν_{L10}^{-1} and ν_{L01}^{-1} respectively. ν_{L10}^{-1} and ν_{L01}^{-1} are updated at iteration *i* using the eqs. (25).

$$\begin{aligned}
\nu_{L10,i}^{-1} &= \frac{1}{\left|S'_L\right|} \sum_{S'_L=\left\{(p_L,q_{L10}) \mid p_L \in S_L,\, \left|l_{p_L}-l_{q_{L10}}\right| \leq \max\left(T_{\Delta l_{LL}},\, 3\nu_{L10,i-1}^{-1}\right)\right\}} \left|l_{p_L}-l_{q_{L10}}\right| \\
\nu_{L01,i}^{-1} &= \frac{1}{\left|S'_L\right|} \sum_{S'_L=\left\{(p_L,q_{L01}) \mid p_L \in S_L,\, \left|l_{p_L}-l_{q_{L01}}\right| \leq \max\left(T_{\Delta l_{LL}},\, 3\nu_{L01,i-1}^{-1}\right)\right\}} \left|l_{p_L}-l_{q_{L01}}\right|
\end{aligned} \tag{25}$$

$S'_L = S_L \setminus s_L$ is comprised of only those sites satisfying the condition (24) for respective neighbors q_{L10} and q_{L01}. s_L is the set of all the sites along the rightmost and topmost borders. Like EM, the proposed constrained ML estimation offers adaptive annealing of the smoothness model parameters by updating them from current disparity map.



(a) (b) I

Figure 3.6: Discontinuity preserving smoothness functions; (a) truncated linear function, (b) exponential function, and (c) Potts model.

3.3.3 Discontinuity handling parameters

Discontinuity in depth occurs at surface boundaries where disparity jumps are bigger than the disparity gradient threshold, i.e. $\left|l_{pL} - l_{qL}\right| > T_{\Delta l_{LL}}$. If such a jump occurs along horizontal line, portions of the image near the boundary in one of the images become invisible in the other. Image points that become invisible in the other image are called occlusions or more specifically half occluded points (since they are invisible in only one of the images).

As mentioned earlier, a truncated linear function is used to handle discontinuity in smoothness. This function is bounded by a minimum value (eq. (26)). This minimum is a threshold parameter which is assumed in graph cuts based algorithms as a value set heuristically. Not complying with statistics, this threshold could be inconsistent for different set of images. In our case, relaxing MRF neighborhood has enabled us to have the threshold parameter explicitly defined as a disparity gradient threshold, therefore, making it consistent for any stereo image pair. Our smoothness function is

$$V\left(l_{pL}, l_{qL}\right) = \min\left(T_{qL}, \left|l_{pL} - l_{qL}\right|\right), \tag{26}$$

where T_{q_L} is the discontinuity handling threshold parameter which equals to $\max\left(T_{\Delta l_{LL}}, 3\nu_{qL}^{-1}\right)$. Discontinuity preserving functions that are frequently observed in the literature are shown in Figure 3.6. These functions are used by graph cuts, belief propagation, and other algorithms. Equation (26) is described by Figure 3.6 (a).

We set T_{q_L}, the depth discontinuity handling parameter, to $\max\left(T_{\Delta l_{LL}}, 3\nu_{L10}^{-1}\right)$ and $\max\left(T_{\Delta l_{LL}}, 3\nu_{L01}^{-1}\right)$ for their respective neighbors. Determination of T_{q_L} for the other neighbors is described in the next section. According to Chebyshev's inequality [Amidan05], multiplying 3 with ν_{L10} is equivalent to setting probability of smoothness between p_L and q_{L10} to $1 - \frac{1}{3^2} \approx .89$; i.e., if $\left|\Delta l_{LL}\right| \le 3\nu_{L10}^{-1}$ then the algorithm assumes that q_{L10} is on the same surface patch p_L is with probability 0.89; otherwise, q_{L10} is an outlier, i.e., it is located on a surface at a different depth. Additionally, if $\left|\Delta l_{LL}\right| \le 1.0$ then

the neighbor is also on the same surface p_L is. Thus, T_{qL} complies with statistics to perform invariably well for any stereo image pair. Occlusions for the data likelihood are handled in a similar fashion by setting T_D to $3\mu_{LL}^{-1}$. Note that T_D also helps match p_L if p_R badly violates the brightness constancy constraint.

3.4 Adaptive support neighborhood

We want to minimize $E_{pL}(l_{pL}) = \sum_{qL \in \mathcal{N}(pL)} \lambda_{D_{LR}} D_{pL}(l_{pL}) + \lambda_{qL} V(l_{pL}, l_{qL})$. Say in the neighborhood, N_H and N_{NH} are the numbers of homogeneous and non-homogeneous points respectively. Then, $|\mathcal{N}(p_L)| = n = n_H + n_{NH}$. Here, we define the homogeneity in terms of color intensities. q_L is homogeneous with p_L if q_L is within a 3*sigma of L_1 distance in intensity value from p_L. Homogeneity is discussed in details later in this section. To understand how homogeneity helps in good matching with proper handling of the discontinuities along the borders, we imagine two cases assuming that convergence in matching of the images has already been achieved. In the first case, we assume that p_L is correctly matched with p_R; then p_R is on the same surface p_L belongs to. Let us approximate the first term in $E_{pL}(l_{pL})$ with zero (since p_L correctly matches). The second term in $E_{pL}(l_{pL})$ is zero if the neighbor q_L is homogeneous with p_L; otherwise, say, second term has the value 3 (i.e. the upper bound as described in Sections 3.3.3 and 3.4.4), since non-homogeneous points are likely to be located on a surface p_L does not belong to. In the second case, we assume that p_L is incorrectly matched; then p_R is likely to be on a surface p_L does not belong to. Then, the first term in $E_{pL}(l_{pL})$ can be approximated with 3 and the second term with 3 for homogeneous points and zero for non-homogeneous points. The two cases are illustrated mathematically in the following.

Case 1: p_L is correctly matched

$$E_{pL}(l_{pL})_{\text{cor}} = n\lambda_{D_{LR}} D_{pL}(l_{pL}) + n_H \lambda_{qL} V(l_{pL}, l_{qL}) + n_{NH} \lambda_{qL} V(l_{pL}, l_{qL})$$
$$= n \times 0 + n_H \times 0 + n_{NH} \times 3 = 3n_{NH} = 3n - 3n_H$$

[by replacing $\lambda_{D_{LR}} D_{pL}(l_{pL})$, $\lambda_{qL} V(l_{pL}, l_{qL})$, and $\lambda_{qL} V(l_{pL}, l_{qL})$ consecutively in the three terms of $E_{pL}(l_{pL})_{\text{cor}}$ by 0, 0, and 3 respectively.]

Case 2: p_L is incorrectly matched

$$E_{pL}(l_{pL})_{\text{inc}} = n\lambda_{D_{LR}} D_{pL}(l_{pL}) + n_H \lambda_{qL} V(l_{pL}, l_{qL}) + n_{NH} \lambda_{qL} V(l_{pL}, l_{qL})$$
$$= n \times 3 + n_H \times 3 + n_{NH} \times 0 = 3n + 3n_H$$

[by replacing $\lambda_{D_{LR}} D_{p_L}(l_{p_L})$, $\lambda_{q_L} V(l_{p_L}, l_{q_L})$, and $\lambda_{q_L} V(l_{p_L}, l_{q_L})$ consecutively in the three terms of $E_{p_L}(l_{p_L})_{\text{inc}}$ by 3, 3, and 0 respectively.]

To obtain a correct match for p_L, we want $E_{p_L}(l_{p_L})_{cor} < E_{p_L}(l_{p_L})_{inc}$ to be true. Clearly, if the number of homogeneous points in $\mathcal{N}(p_L)$ is increased, the probability of the condition, $E_{p_L}(l_{p_L})_{cor} < E_{p_L}(l_{p_L})_{inc}$, to be satisfied becomes high. In building a suitable $\mathcal{N}(p_L)$, we, therefore, apply homogeneity. Our approach is the following. First, we define a minimum size window and a maximum size window. We include all the points that are in the minimum size window into $\mathcal{N}(p_L)$. Then, we pick the points that are homogeneous with p_L from the points inside the maximum size window but they are outside the minimum size window (see Figure 3.7).

3.4.1 Homogeneous neighbors selection

We use the data parameter μ_{LL} to pick the homogeneous neighbors. First, we estimate the mean intensity, $\bar{I}(p_L)$, of the site p_L using mean shift algorithm. q_L is homogeneous with p_L; therefore, $\bar{I}(p_L)$ is also the mean intensity of q_L. p_L is homogeneous with q_L if the intensity difference between p_L and q_L is only due to noise. The variables $\bar{I}(p_L) - I(p_L)$ and $\bar{I}(p_L) - I(q_L)$ are identical, but they also are independent. In consequence, applying variance analysis, we find

$$Var\left(\left|\bar{I}(p_L) - I(p_L)\right|\right) = \frac{1}{2}\left\{Var\left(\left|I(p_L) - I(q_L)\right|\right) + \left(E\left(\left|I(p_L) - I(q_L)\right|\right)\right)^2\right\},$$

where $Var\left(\left|I(p_L) - I(q_L)\right|\right) = \left(E\left(\left|I(p_L) - I(q_L)\right|\right)\right)^2 = \mu_{LL}^{-2}$, since $\left|I(p_L) - I(q_L)\right|$ is exponentially distributed. A homogeneous neighbor is defined to be within a range of $3\mu_{LL}^{-1}$ (according to the 3*sigma confidence) from $\bar{I}(p_L)$. Accordingly, we have the condition in (27).

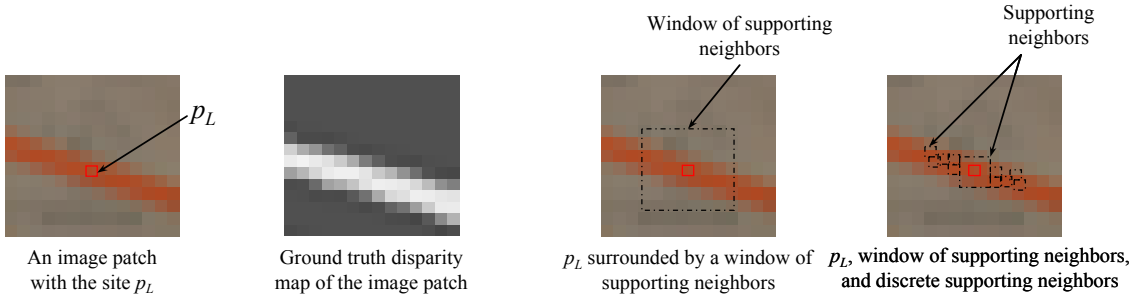
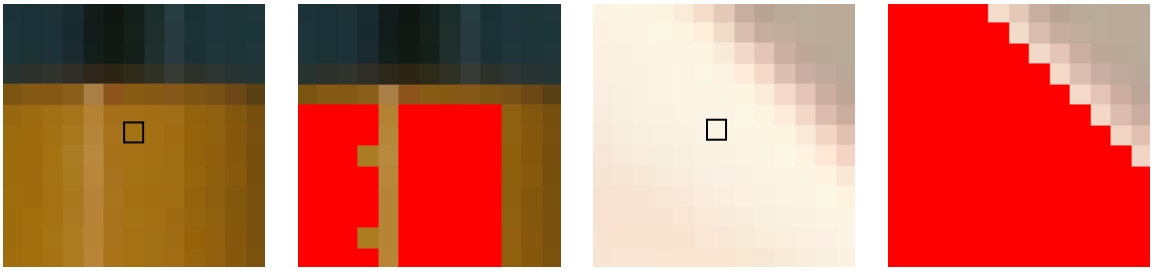


Figure 3.7: Demonstration of role of homogeneity for discontinuity handling



(a) (b) I (d)

Figure 3.8: (a) Image patch 1 with center pixel shown with a black box as point *p*, (b) homogeneous points picked by the mean shift algorithm; (c) Image patch 2 and (d) the homogeneous points.

To observe the performance of the parameter μ_{LL}^{-1} as a threshold in picking points that are in color homogeneous with p_L, we conduct a few experiments. First, we find the mean intensity, $\bar{I}(p_L)$, of the site p_L using the mean shift algorithm; we describe how to find $\bar{I}(p_L)$ later. Then, we pick the points that are within a range of $3\mu_{LL}^{-1}$ from the mean intensity (inequality (27)).

$$\left|\bar{I}(p_L) - I(q_L)\right| \leq 3\mu_{LL}^{-1}. \tag{27}$$

Results from the experiments are shown in Figure 3.8. Two image-patches in (a) and (b) are taken from the image 'Tsukuba'. Centers of the patches shown in black rectangles are p_L.

All q_L within the minimum size window and those individually picked outside the window but inside the maximum size window comprise the neighborhood, $\mathcal{N}(p_L)$. Figure 3.9 shows neighborhood sizes for each pixel in two images – 'Tsukuba' and 'Venus'. Brightness of a point p_L in the images in the figure is proportional to the number of neighbors in $\mathcal{N}(p_L)$. As described by the images, size of $\mathcal{N}(p_L)$ is big in homogeneous regions and small near the borders with high texture variations.

3.4.2 Summary of the adaptive support neighborhood selection

Say that the mean intensity of p_L is $\bar{I}(p_L)$. An estimate of the reference intensity is the average of the intensities of the neighboring pixels that are similar to p_L in color. In

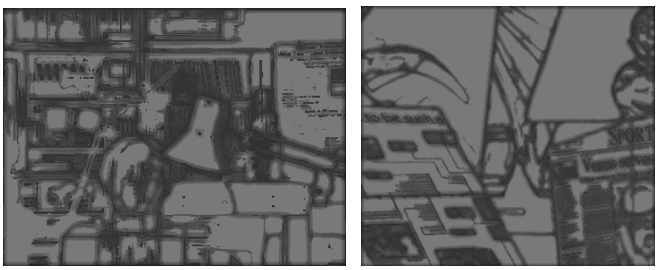
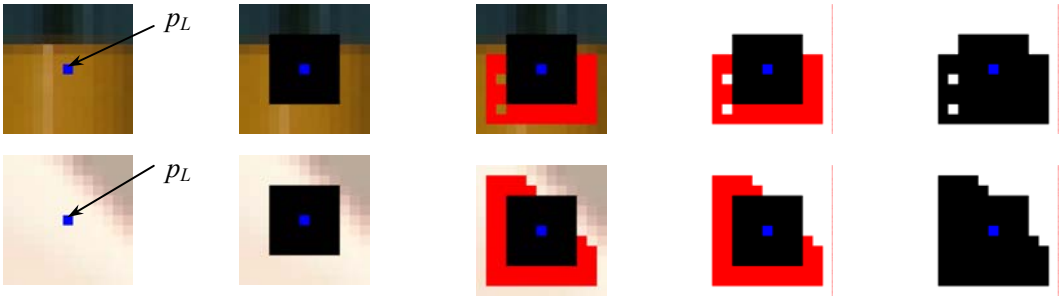


Figure 3.9: Neighborhood sizes for 'Tsukuba' (left) and 'Venus' (right) obtained applying homogeneity; brightness is proportional to the neighborhood size.



(a) (b) (c) (d) (e)

Figure 3.10: Graphical representations of selection of $\mathcal{N}(p_L)$; (a) point p_L and its' surrounding, (b) select a window, $W_{\min}$, (c) pick the points homogeneous with p and outside $W_{\min}$, but inside $W_{\max}$, (d) all the points selected inside the $W_{\max}$, and (e) the adapted window and its' points

the next subsection we describe how $\bar{I}(p_L)$ is estimated using mean shift algorithm. Say that $W_{\max}$ and $W_{\min}$ are the maximum and minimum windows respectively. Then, the steps for selection of the neighborhood $\mathcal{N}(p_L)$ are

1. Select the window $W_{\min}$ centering p_L
2. Include all the points of $W_{\min}$ in $\mathcal{N}(p_L)$
3. Pick the points q_L that are homogeneous with p_L and also inside the $W_{\max}$ window, but they are outside $W_{\min}$; add these q_L to $\mathcal{N}(p_L)$

Figure 3.10 shows each of the neighborhood selection steps in pictorials. In Figure 3.10(a), we show image patches centered at p_L (the blue point at the center). Figure 3.10(b) shows the minimum size window $W_{\min}$ (in black including p_L). The points outside $W_{\min}$ that satisfy the condition in (27) are shown in red in Figure 3.10(c). Figure 3.10(d) shows all the points that are included in $\mathcal{N}(p_L)$. Figure 3.10(e) shows all the neighbors in black and the point p_L in blue, which is also included in $\mathcal{N}(p_L)$.

For all experiments, we hold $W_{\min}$=7×7 and $W_{\max}$=11×11. Notice that $W_{\max}$ does not need to assume a strict boundary; rather, a boundary is defined only to pick enough number of homogeneous points so that along the surface boundaries $E_{pL}(l_{pL})_{cor} < E_{pL}(l_{pL})_{inc}$ has a high probability to satisfy.

3.4.3 Mean shift algorithm to determine the reference intensity

$\bar{I}(p_L)$ is the mean intensity of p_L, we want to estimate $\bar{I}(p_L)$ with the following mean shift algorithm:

2. Select the window $W_{\max}$ around p_L ; say, the time stamp, $t \leftarrow 0$

3. Select the intensity of p_L as the initial value for $\bar{I}(p_L)_t = I(p_L)$
4. Pick the points q_L that satisfies the inequality (27); say, these q_L form $\mathcal{N}(p_L)_t$; apply $t \leftarrow t+1$
5. Compute $\bar{I}(p_L)_t \leftarrow \frac{1}{\left|\mathcal{N}(p_L)_t\right|} \sum_{q_L \in \mathcal{N}(p_L)_t, \left|\bar{I}(p_L)_{t-1} - I(q_L)\right| \leq 3\mu_{LL}^{-1}} I(q_L)$
6. Repeat steps 3 and 4 until $\left|\bar{I}(p_L)_t - \bar{I}(p_L)_{t-1}\right| \leq T$ is satisfied; here, T is a small threshold value which can be set to zero and at the same time can be tied with a empirically chosen as fixed but sufficiently big number of iterations.

The mean shift algorithm described above does not assume a moving window as described by Cominiciu and Meer in [Comaniciu02]; rather the window remains at the same location and the mean is updated relying on the data points residing only within the window.

3.4.4 Smoothness parameters for the supporting neighbors

So far, we have introduced an adaptive support neighborhood based technique to find the minimum of the aggregated energy, $\sum_{p_L \in S_L} E_{p_L}(l_{p_L})$. The key idea in this neighborhood support technique is building confidence in matching by applying eq. (15) on many neighbors in the support neighborhood. We introduce this support neighborhood in the cost function in eq. (15) to achieve global convergence in matching and at the same time handle discontinuity properly. Now we describe how smoothness model parameters for each point in $\mathcal{N}(p_L)$ are obtained from the two basic smoothness model parameters, ν_{L10}^{-2} and ν_{L01}^{-2}, introduced and estimated earlier. Computing n number of smoothness parameters for *n* number of neighbors is computationally expensive. Instead, we apply an observation that variances are linearly related, i.e., as the Euclidean distance of q_L from p_L, which we call d(p_L, q_L), increases, the variance of their smoothness model variable increases linearly (see Figure 3.11). The observation is validated from the Middlebury ground truth data. We estimate the variances of Δl_{LL} at d(p_L, q_L)=1,…,5 along horizontal and vertical directions and plot them (see Figure 3.11). It is observed from the plots that the variance increases almost linearly with d(p_L, q_L). This observation can be explained by the fact that image patches are locally almost planar. As d(p_L, q_L) increases, Δl_{LL} increases linearly at a rate of the slope of the planar patch along a particular direction.

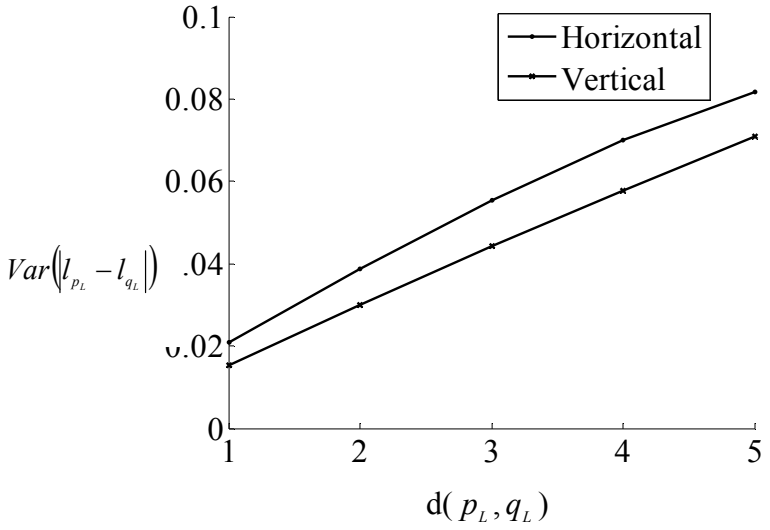


Figure 3.11: Variance of disparity gradient, $Var\left(\left|l_{p_L} - l_{q_L}\right|\right)$, vs d(p_L , q_L) along horizontal and vertical directions for the test image, 'Tsukuba'

Now, say that p_L is located at $\left(x(p_L), y(p_L)\right)$ and its neighbor q_L at $\left(x(q_L), y(q_L)\right)$. Applying our observation, we infer variances for neighbor q_L from ν_{L10}^{-2} and ν_{L01}^{-2} according to the following formula (see also Figure 3.12),

$$\nu_{q_L}^{-2} = \left|x(q_L) - x(p_L)\right|\nu_{L10}^{-2} + \left|y(q_L) - y(p_L)\right|\nu_{L01}^{-2}, \tag{28}$$

which is a first order polynomial that performs mapping of the prior $\nu_{q_L}^{-2}$ for q_L. Variances are sign invariants; therefore, we use absolute value for both axes. In consequence, the mapping is quadrant-wise planar. Similar to q_{L10} and q_{L01}, the discontinuity handling parameter T_{q_L} for any neighbor q_L in the supporting neighborhood is set to $\max\left(T_{\Delta l_{LR}}, 3\nu_{q_L}^{-1}\right)$.

Mapping prior model parameters for the neighbors in the supporting neighborhood according to eq. (28) is remotely related to existing works developed for adaptive window based stereo matching algorithms. Several of these algorithms are described and a new one is proposed in [Kanade94] by Kanade and Okutomi, where, disparity gradient thresholds are either both defined manually and kept constant for all the window members or dependent on only on the distances of the points from the candidate point along x-axis, except in the method described by Kanade and Okutomi, which is statistical and takes Euclidean distance into account. These algorithms are for determining adaptive windows that provide more accuracy in matching than algorithms having windows not adapted. We use an adaptive support neighborhood technique for handling discontinuity in smoothness in MRF. One important difference of our approach from algorithms

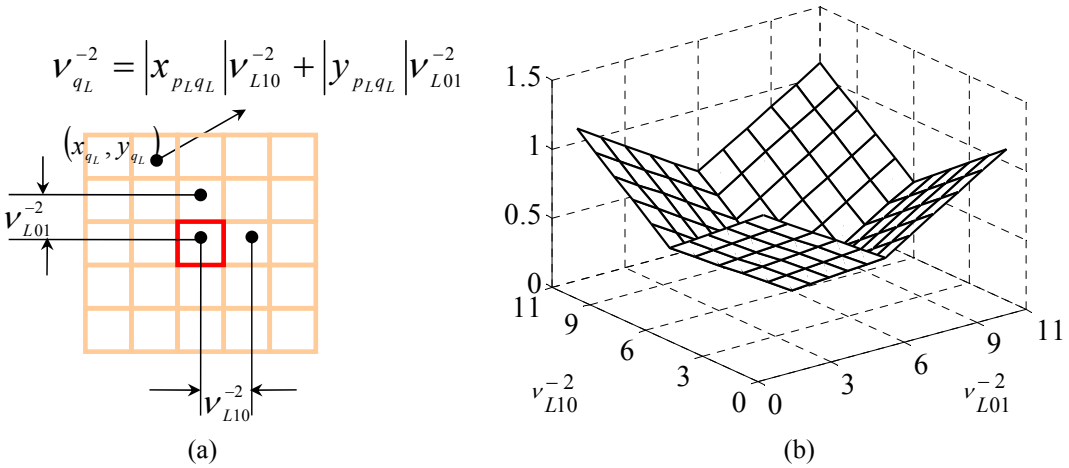


Figure 3.12: (a) Local Support Neighborhood and (b) Geometric representation of estimation of $\nu_{q_L}^{-2}$

described by [Kanade94] is that we do not estimate the parameters for each member in the neighborhood; rather, we obtain them from two basic parameters for the two neighbors, q_{L10} and q_{L01}. We find a linear relationship among the parameters and consequently, we are able to consider the Euclidean distances of the window members from *p* without estimating the parameters individually for all the neighbors.

3.5 Summary of the proposed algorithm

In this section, first, we summarize the cost function and its parameters; the purpose is to have a look on the equations and their parameters altogether and describe the matching algorithm in a flow diagram for clarity. Then, we describe how the images are matched symmetrically for improved performance and also to detect occlusions.

3.5.1 Matching the stereo pair

The algorithm picks a site p_L from the left image, S_L, and finds its minimum cost in a winner take all (WTA) fashion. Following are listed the cost function and all its parameters.

$$E_{p_L}(l_{p_L}) = \sum_{q_L \in \mathcal{N}(p_L)} \lambda_D D_{p_L}(l_{p_L}) + \lambda_{q_L} V(l_{p_L}, l_{q_L}), \text{ where, } \lambda_D = \mu_{LL} \text{ and } \lambda_{q_L} = \nu_{q_L}.$$

$$D_{p_L}(l_{p_L}) = \min(T_D, |I(p_L) - I(p_R)|), \text{ where } T_D = 3\mu_{LL}^{-1}$$

$$V(l_{p_L}, l_{q_L}) = \min(T_{q_L}, |l_{p_L} - l_{q_L}|), \text{ where, } T_{q_L} = \max(T_{\Delta l_{LL}}, 3\nu_{q_L}^{-1}) \text{ and } T_{\Delta l_{LL}} = 1.0$$

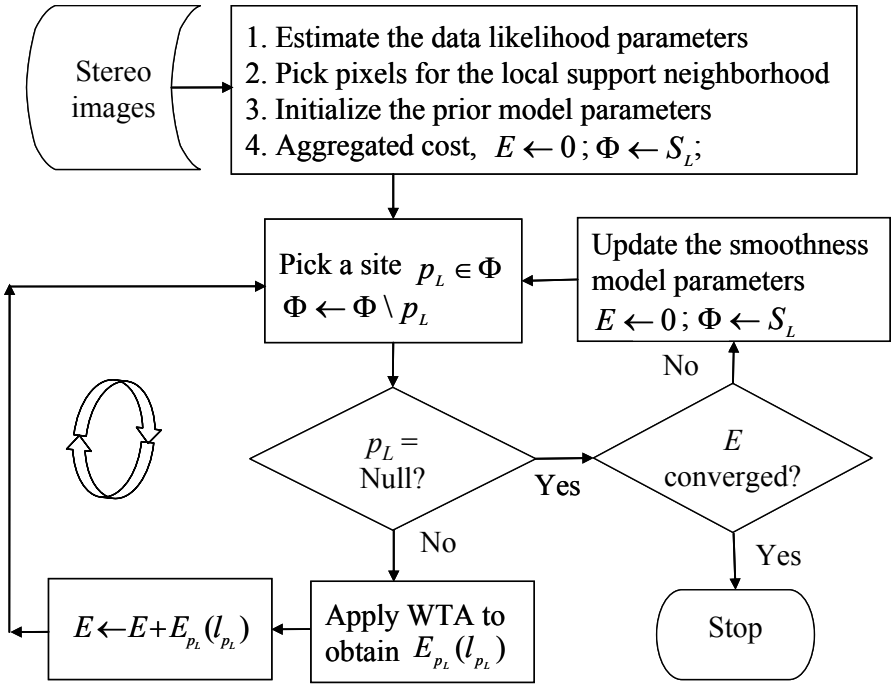


Figure 3.13: Flow diagram of the proposed algorithm

We find minimum cost of a point using WTA as described in the following,

$$l_{p_L} = \underset{l_{pL} \in \{d_{\min},\ldots,d_{\max}\}}{arg\min} \sum_{q_L \in \mathcal{N}(p_L)} \lambda_D D_{p_L}(l_{p_L}) + \lambda_{q_L} V(l_{p_L}, l_{q_L}).$$

A flow diagram of the matching algorithm is shown in Figure 3.13. First, we initialize the system: we estimate the data likelihood model parameters from one of the images and initialize the smoothness model parameters v_{qL} with small values. Then, we start our alternating optimization algorithm which has two phases. In the first phase, we update the prior model parameters; in the second, we pick the candidate sites p_L one by one from the left image and match in the right. When matching of all the sites is done, we come back to the first phase and let the iteration continue similar way until convergence in the aggregated cost is achieved. The convergence is achieved when difference of aggregated costs from two consecutive iterations is less than a small number predefined by the user. Alternatively, or at the same time, the iterations can be tied to an empirically determined sufficiently big number to avoid infinite looping.

3.5.2 Symmetric matching and occlusion detection

The flow diagram in Figure 3.13 is for only one way matching. We develop a pair of symmetric cost functions and match the images symmetrically to reduce the matching error by about 1% averaged on all the test image pairs. In symmetric matching, the smoothness energy is computed from both the stereo images. First, we perform the left to

right matching. In one way matching, we assumed that disparities of only the neighbors, q_L, are given. In symmetric matching, additionally we consider that disparities of q_R are also given. Here, q_R in the right image are the correspondences of q_L. Thus, the inference algorithm uses disparity information from both images. In the same way, we perform the right to left matching. The symmetric matching equations are described in equation set (29) with the first equation for the left to right match and the second for right to left.

$$\begin{aligned} E_{p_L}(l_{p_L}) &= \sum_{q_L \in \mathcal{N}(p_L)} \lambda_D D_{p_L}(l_{p_L}) + \lambda_{q_L}\left(\frac{1}{2}V_{LR}(l_{p_L}, l_{q_L}) + \frac{1}{2}V_{RL}(l_{p_R}, l_{q_R})\right), \text{ where } p_L \in S_L \\ E_{p_R}(l_{p_R}) &= \sum_{q_R \in \mathcal{N}(p_R)} \lambda_D D_{p_R}(l_{p_R}) + \lambda_{q_R}\left(\frac{1}{2}V_{RL}(l_{p_R}, l_{q_R}) + \frac{1}{2}V_{LR}(l_{p_L}, l_{q_L})\right), \text{ where } p_L \in S_R \end{aligned} \tag{29}$$

S_R is the right image. Since both sets of corresponding neighbors are equally likely, each of the smoothness terms are multiplied by half. Due to uniqueness constraint, we have only one data term in both the equations. In the symmetric matching algorithm, first we perform the left to right and then right to left matching. Then we come back to the left to right match and the iteration continues similar way.

4 Occlusion filling in stereo

Regions of one of the stereo images that are invisible in the other are called occlusions or more specifically half occlusions. Occlusions occur near the image borders. They also always appear inside the images when two or more distinct surfaces appear as foregrounds and backgrounds in the scene. The border occlusion occurs due to the right camera being located at the right side of the left camera and thus missing some of the left portion of the field of view of the left camera. Inside the scene, part of the background near the border of the two surfaces becomes invisible (see Figure 4.1). We call these occlusions the non-border occlusions.

The non-border occlusions can be of three types: partial occlusions, self occlusions, and total occlusions (Figure 4.2). In partial occlusions, only part of a visible background surface becomes invisible to the right camera. In self occlusions, only part of the visible foreground surface which is rounded becomes invisible to the right camera. If the rounded surface is also lacking gradient at the same time, then the surface appears planar to the binocular stereo, making detection of the occlusion impossible. In total occlusion, an isolated surface is visible to the left camera, but the isolated surface remains entirely invisible to the right camera. Therefore, the surface can not be interpolated from non-occluded surface. The three non-border occlusion creation processes are described by Figure 4.3. Figure 4.4 shows disparity maps with border and non-border occlusions in the middle column.

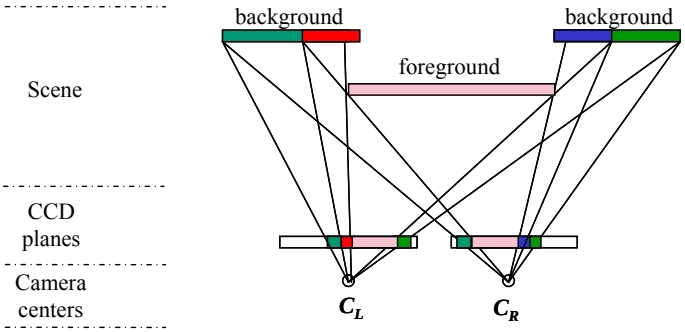


Figure 4.1: Demonstration of creation of occlusions

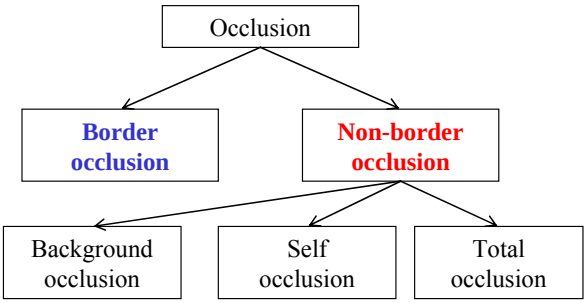


Figure 4.2: Classification of occlusions

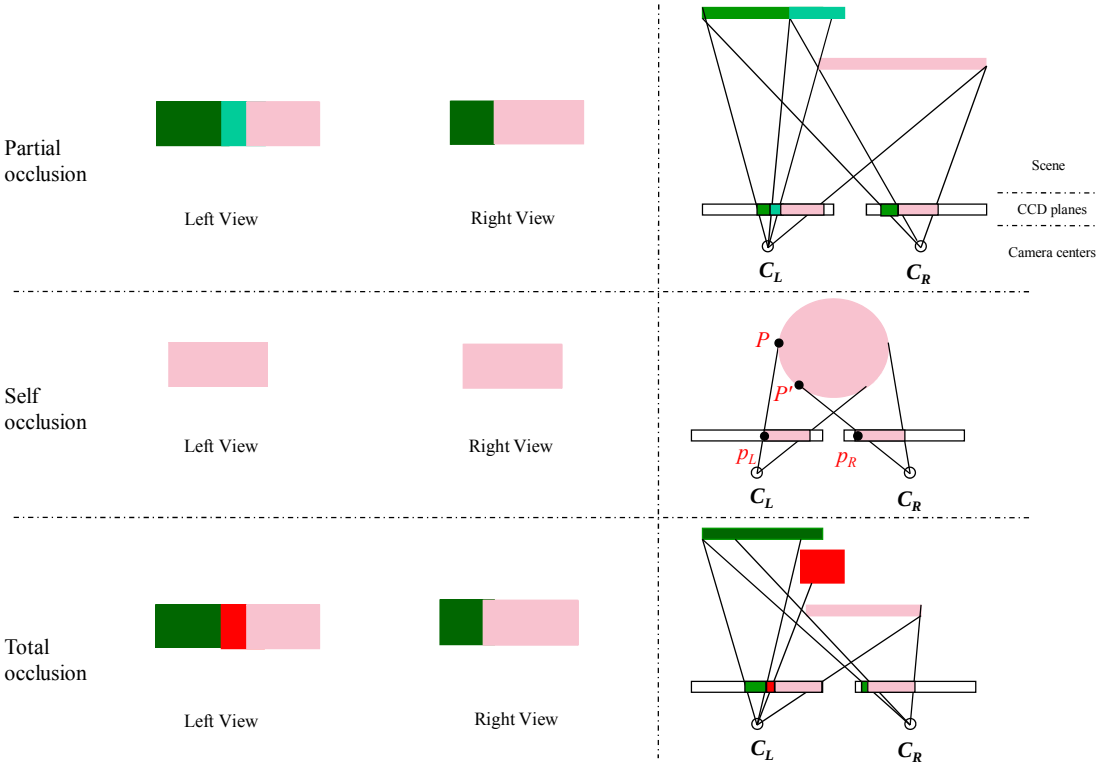


Figure 4.3: Left column – Views of 3D scene as seen by the left and right cameras, right column – cross-section of the top view of scene and camera CCD planes set up

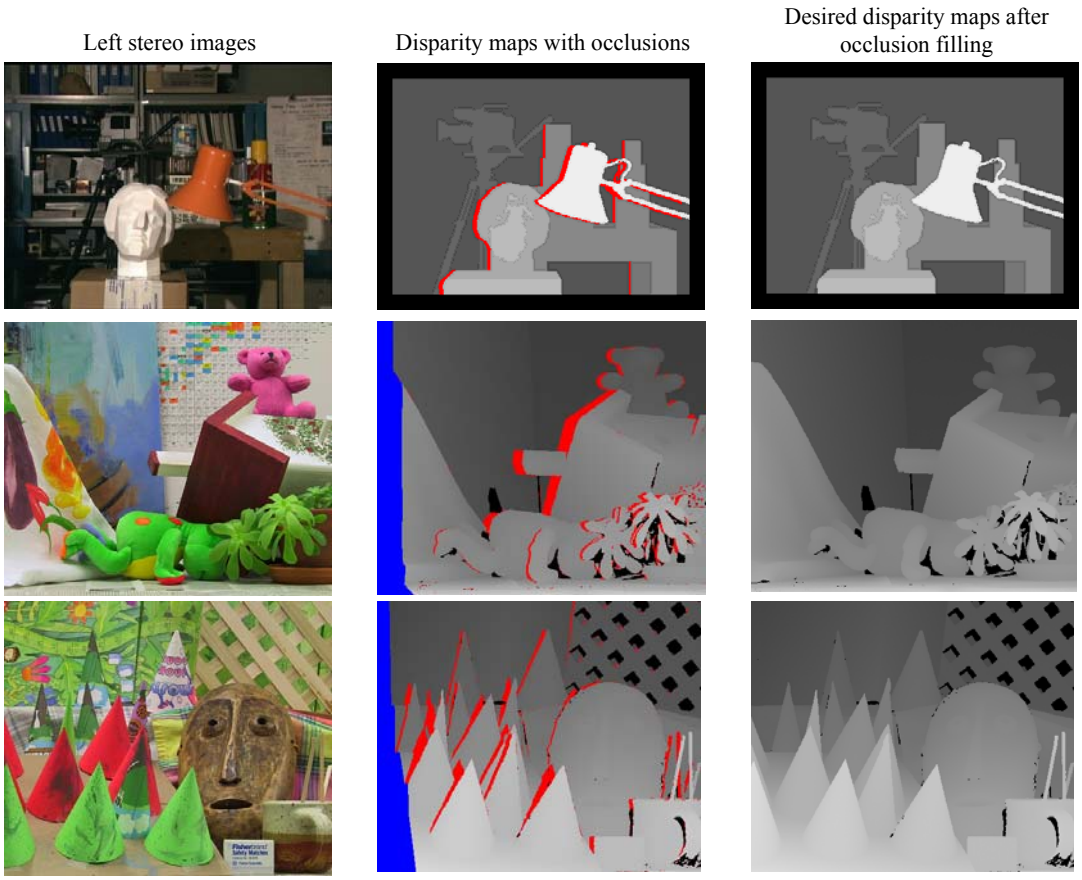


Figure 4.4: Border occlusions (blue) and non-border occlusions (red) in some of the Middlebury ground truth disparity maps

Conventional stereo vision algorithms have not given much attention on occlusion filling. The main focus of many stereo vision algorithms is the matching part, although, occlusion filled 3D models are highly expected in many rendering applications. Due to the importance of occlusion filling, Middlebury College has emphasized on an evaluation of matching performance for all pixels that include both occluded and non-occluded pixels of the disparity maps.

In Chapter 2, in a literature review on occlusion filling, we have discussed that occlusion filling algorithms are either not independent of the matching algorithm or they do not account for the slope of surrounding surface patch. Most of the times, occluded regions extend with the slope of nearby non-occluded regions. To incorporate slope information in estimation of disparity of an occluded point it is necessary to perform extrapolation of the disparity of the occluded point from its non-occluded neighbors. The goal of the study of occlusion filling algorithms in this chapter is to deliver an occlusion filling algorithm that is independent of the matching algorithm and also accounts for the surface slope.

In the remaining of this chapter, we present theory of a number of existing and proposed occlusion filling algorithms. We call one of these existing algorithms neighbor's disparity assignment (NDA) that directly assigns disparity to the occluded point reading the disparity from one of its non-occluded neighbors. Another algorithm described is called diffusion in intensity space (DIS), which is inspired by the paper, [Min08]. We propose two new algorithms: weighted least squares (WLS) and segmentation based least squares (SLS). Depending on the location of the occluded point and the relative locations of its neighbors, we need to perform extrapolation or interpolation to account for the slope of the surrounding surface patch. We assume that a surface patch, comprised of the neighbors of an occluded point, is planar. Linear model works as a prior on the disparity map. In a small neighborhood, disparity map could be wrongfully non-linear for two regions: disparities of some of the neighbors are inaccurate or there exist stair effects due to discrete disparity values. A linear model enforces planarity in a small neighborhood and thus minimizes wrongfully introduced non-linearity in the neighborhood. In our proposed algorithms, least squares estimation technique is used to estimate parameters of this plane. It is known that theoretically, interpolation and extrapolation both can be applied with least squares to estimate the unknown value if the underlying model is already known. Both of these techniques apply depending on relative locations of the non-occluded neighbors. For instance, in filling border occlusions the technique should always be extrapolation, since the non-occluded neighbors are always in one side (right) of the occluded point. With respect to our least squares estimation technique both interpolation and extrapolation are equivalent. Therefore, we can refer to both of these techniques as either of them. By choice, we will use the term interpolation hereafter without worrying about the relative locations of the neighbors.

4.1 Neighbor's Disparity Assignment (NDA)

Figure 4.5 shows a simple flow diagram of an existing occlusion filling algorithm called neighbor's disparity assignment (NDA). In this algorithm, the border occlusions are filled with the disparities of the non-occluded points located at the right side and the non-border occlusions are filled with the disparities of the non-occluded points located at the left side. NDA assumes that the surface patch surrounding the occluded point is fronto-parallel. We conducted experiments with NDA for the ground truth disparity maps and found matching error in the non-occluded regions as listed in Table 4.1. Figure 4.6 shows some of the occlusions filling results of these experiments visually. It is evident from the results that many occlusion regions are incorrectly filled. In the figures, the invalid fillings are marked with circles.

4.2 Weighted Least Squares (WLS)

In weighted least squares (WLS) approach, all non-occluded neighbors are considered as valid neighbors in a neighborhood (see Figure 4.7). Valid neighbors work as control points in the interpolation. Similar to the order in NDA, the border occlusions are filled in the right-to-left direction and the non-border occlusions are filled in the opposite order. We have assumed linear model for the interpolation; however, neighbors on the foreground introduce non-linearity. Usually, points on the foreground of a scene have distinctive color intensities compared to the background. Applying this cue, we set the values of the weights such a way that the weights suppress influence of the foreground points in the interpolation.

In a weighted least squares approach, each residual error term in the aggregated residual is weighted. Thus, the aggregated residual is defined as

$$\Delta = \sum_{q_L \in \mathcal{N}(p_L)} w_{q_L} \left(\hat{I}_{p_L}(q_L) - I_{p_L}(q_L) \right)^2 , \tag{30}$$

where $w_{qL} = \exp\left(- \frac{\left| \bar{I}(p_L) - I(q_L) \right|}{\mu_L} \right)$ is chosen heuristically as the likelihood of p_L with its neighbor q_L. The heuristic choice for likelihood comes from the fact that if a neighbor has a small $\left| \bar{I}(p_L) - I(q_L) \right|$ value, the neighbor q_L is located most likely in the same surface p_L is located on, and thus, linearity is enforced. It is a strong smoothness property in natural images that neighboring pixels with similar color intensities tend to live on the same surface patch. Δ is partially differentiated with respect to parameters of the unknown linear model to obtain a system of linear equations. We solve these equations with least squares.

Say that $\mathbf{F} = \begin{bmatrix} x_1 & y_1 & 1 \\ \vdots & \ddots & \vdots \\ x_N & y_N & 1 \end{bmatrix}$ is the matrix of the coordinates of all the control points (non-occluded neighbors) and $\mathbf{L} = \begin{bmatrix} l_1 & l_2 & \dots & l_N \end{bmatrix}'$ is the vector of the corresponding labels. Then the linear model is

$$l_{pL} = a + bx(p_L) + cy(p_L), \tag{31}$$

where $(x(p_L), y(p_L))$ is the coordinate of p_L; a, b, and c are the model parameters. Also, say that the weight vector corresponding to the control points matrix is $\mathbf{w} = \begin{bmatrix} w_{q_{L1}} & w_{q_{L2}} & \dots & w_{q_{LN}} \end{bmatrix}'$. We compute $\mathbf{F}_\mathbf{w} = diag(\mathbf{w})\mathbf{F}$ and $\mathbf{L}_\mathbf{w} = diag(\mathbf{w})\mathbf{L}$. If $\mathbf{P} = \begin{bmatrix} a & b & c \end{bmatrix}'$ is the parameter vector of the linear mapping,

$$\mathbf{P} = (\mathbf{F}_\mathbf{w}^\mathsf{T}\mathbf{F}_\mathbf{w})^{-1}\mathbf{F}_\mathbf{w}^\mathsf{T}\mathbf{L}_\mathbf{w} \tag{32}$$

Once $\mathbf{P}$ is estimated, disparity of the occluded point p_L is estimated as

$$\hat{l}_{p_L} = \begin{bmatrix} 1 & x(p_L) & y(p_L) \end{bmatrix}\mathbf{P}. \tag{33}$$

4.3 Diffusion in Intensity Space (DIS)

This method is inspired by a recently published paper, [Min08]. In the paper, Min and Sohn have solved stereo matching iteratively with non-linear diffusion. The authors estimated weighted diffusion energy which indicated a good match at its lowest value at a particular disparity label. After detecting occlusions from left-to-right and right-to-left disparity maps they approximated diffusion energy of the occluded region to determine disparities of the occluded points. One main disadvantage with this algorithm is that the occlusion filling algorithm is not independent to the stereo matching algorithm. The energies in the last iteration of the matching are considered as the initial diffusion energies of the occlusion filling iterations. Besides, estimation of diffusion energy does not include interpolation mechanism, i.e., disparity plane of occluded regions is not able to maintain the slope of the disparity plane of surrounding non-occluded regions. Yet we wanted to study performance of this algorithm by incorporating its concept of diffusion in intensity space in a proposed algorithm. In our proposed algorithm, called DIS, the diffusion energy $E(p_L)$ is initially assigned with zero when p_L is non-occluded.

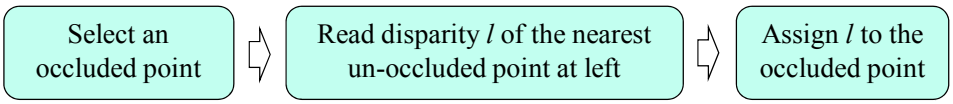


Figure 4.5: Flow diagram of existing neighbor disparity assignment approach

Diffusion energies of the non-border and border occlusions points are updated according to the equations (34) and (35) respectively.

$$E(p_L)=\min_{l_{pL}=\{0,\ldots,l_{\max}\}}\left(\frac{1}{2\left|q_L\in\mathcal{N}(p_L)\wedge l_{q_L=l_{pL}}\right|}\sum_{q_L\in\mathcal{N}(p_L)\wedge l_{qL=l_{pL}}}\left(\left|\bar{I}(p_L)-\bar{I}(q_L)\right|+E(q_L)\right)\right) \tag{34}$$

$$E(p_L)=\min_{l_{pL}=\{0,\ldots,l_{pLf}-2\}}\left(\frac{1}{2\left|q_L\in\mathcal{N}(p_L)\wedge l_{q_L=l_{pL}}\right|}\sum_{q_L\in\mathcal{N}(p_L)\wedge l_{qL=l_{pL}}}\left(\left|\bar{I}(p_L)-\bar{I}(q_L)\right|+E(q_L)\right)\right) \tag{35}$$

The update equations basically integrate the energies of non-occluded points having the same disparity label. Non-occluded point p_L is assigned to the disparity that corresponds to the minimum $E(p_L)$ estimated for all possible disparities. For non-border occlusions the minimum $E(p_L)$ is taken over the range from 0 to $l_{p_{Lf}}$. Here, $l_{p_{Lf}}$ is disparity of the non-occluded point located at the right side of p_L; this non-occluded point belongs to the foreground which has disparity bigger than the disparity of p_L. The border occlusions are filled in the right-to-left direction and the non-border occlusions are filled in the left-to-right direction. Pseudo codes of DIS algorithm are presented in Figure 4.8.

4.4 Segmentation based Least Squares (SLS)

SLS is our last proposed algorithm. One major difference between SLS and WLS is that in SLS, the control points are rather a subset of the neighbors. The control points are segmented out from the neighborhood by applying visibility constraint, disparity gradient constraint, and color similarity to be detailed shortly in this section. Interpolation accounts for the slope of the non-occluded surface in the neighborhood. An occluded region could be part of a slanted surface requiring the assigned disparities to be interpolated accordingly. Thus, our approach amounts to the following sequence of operations: picking an occluded point, selecting control points from the neighborhood of the occluded point, and then interpolating disparity of the occluded point from the control points (Figure 4.9).

Table 4.1: Performance (percentages of matching error) of neighbor assignment

Tsukuba	Venus	Cones	Teddy	Map	Sawtooth
7.8%	5.5%	32.1%	24.6%	4.7%	10.2%

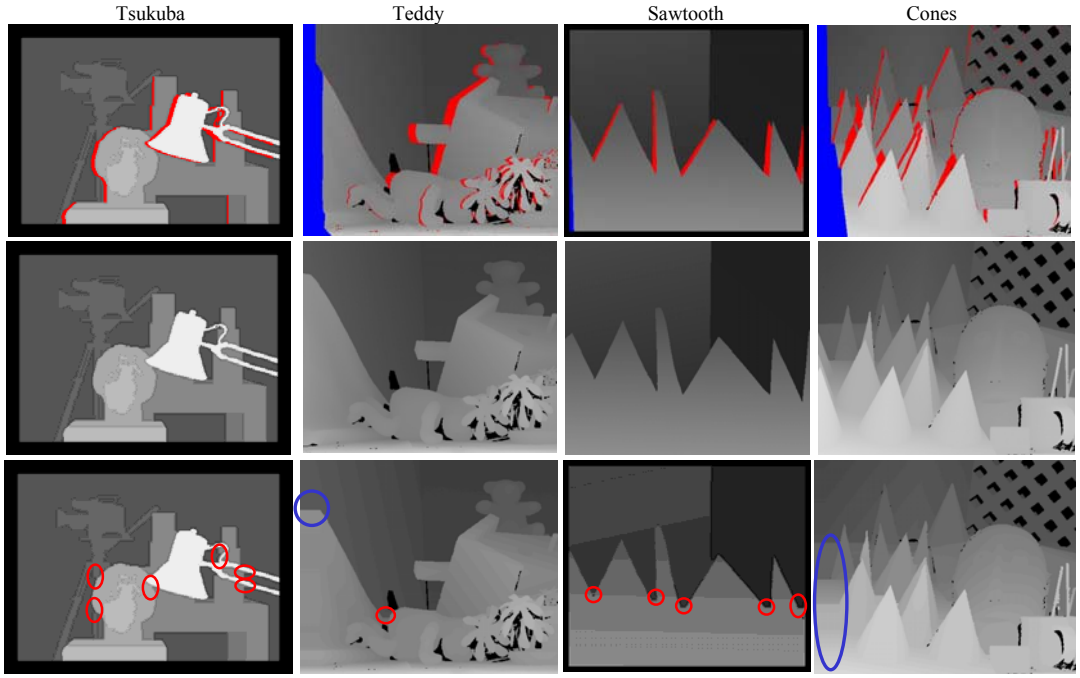


Figure 4.6: Top – ground truth disparity maps with occlusions, middle- ground truth disparity maps, bottom – occlusion filling with NDA (wrong fillings are marked with circles)

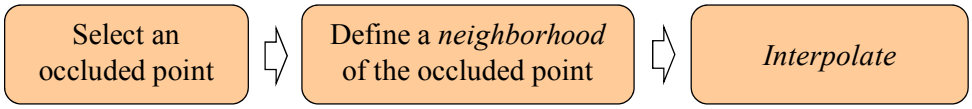


Figure 4.7: Flow diagram of WLS approach for occlusion filling

ALGORITHM: Occlusion filling with diffusion

Initialize the diffusion energies: $E(p_L) = 0$

For *all occluded points*

Estimate $E(p_L)$
l_{p_L} *= disparity corresponding to minimum* $E(p_L)$
Mark p_L *as non-occluded*

End For

Note: Border and non-border occlusions are filled in the right-to-left and left-to-right directions respectively.

Figure 4.8: Pseudo codes for DIS

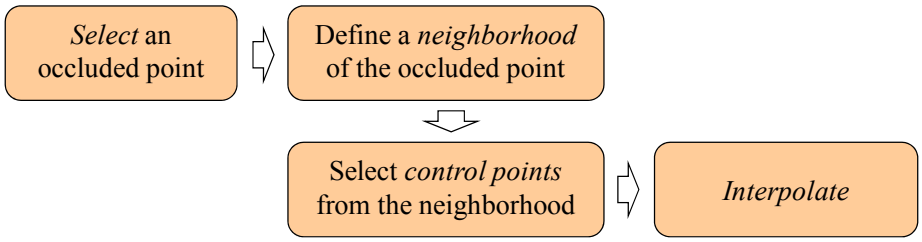


Figure 4.9: A basic flow diagram of the proposed SLS occlusion filling algorithm

Selection of the control points need to be done from a group of potential neighbors (Figure 4.10). Difficulty arises when the potential neighbors come form multiple background surfaces, since the occluded point is located on only one of the surfaces (see Figure 4.11). When there are narrow objects in the scene, the condition that the occluded points have smaller disparities than the disparities of the non-occluded points in the right is violated (Figure 4.12).

Another issue considered in SLS occlusion filling is the order in filling. When an occluded region contains more than one surface, an occluded point may not at all belong to the surface formed by its non-occluded neighbors; rather, the occluded point may be located on a surface formed only by the occluded points in its neighborhood (Figure 4.13), which need to be interpolated first. Occluded points that are attached to at least one non-occluded point are considered as potential points for filling. Among these potential points, the occluded point which has the lowest *homogeneity* estimate (i.e., the point exhibits high homogeneity or high color similarity with its neighbors) is filled first. An occluded point has low homogeneity estimate if the number of neighbors with color intensity similar to its own color intensity is also high. Thus, we use color intensity information and apply homogeneity estimate to find order priority in filling.

4.4.1 Theory of segmentation and filling order

Say that p_L is the occluded point we want to fill and $\mathcal{N}(p_L)$ is defined as a set of non-occluded neighbors of p_L. Our goal is to find the valid control points that participate in the interpolation. Say that $\mathcal{N}(p_L)$ is currently empty. According to the occlusion formation process, disparity of occluded point is less than the disparity of the foreground. A foreground point can be extracted by picking the nearest non-occluded point at the right side of p_L. We call this point and its disparity p_{Lf} and $l_{p_{Lf}}$ respectively. The non-occluded neighboring points that have disparities less than $l_{p_{Lf}}$ are taken into $\mathcal{N}(p_L)$. If the occlusion is created by a foreground, which is a narrow object, the non-occluded points in both sides (left and right) of p_L bear similar disparities. Therefore, we define a second condition based upon the disparity gradient constraint. Say, the nearest non-occluded point in the left side is p_{Lb} and its disparity is $l_{p_{Lb}}$. We apply the disparity gradient constraint $\left|l_{p_{Lb}} - l_{q_L}\right| \leq 1$ for a neighbor q_L with disparity l_{q_L} and include the

neighbors satisfying the constraint, in $\mathcal{N}(p_L)$. The combined condition is described in (36).

$$\left|l_{pLb} - l_{qL}\right| \leq 1 \vee l_{qL} < l_{pLf} \tag{36}$$

While $\mathcal{N}(p_L)$ is formed, the points q_L in $\mathcal{N}(p_L)$ may come from two or more background surfaces. If we observe the ground truth disparity maps of the Middlebury test images, we find that in a small neighborhood, more than three surfaces are not present. Such neighborhoods may exist but in a negligibly low number. Therefore, we can *safely* assume that $\mathcal{N}(p_L)$ contain points from not more than two surfaces. These two surfaces have points with disparities in two distinct ranges and one of these two surfaces is closer to the camera than the other. Therefore, the minimum $l_{\min}$ of all the disparities of the pixels in $\mathcal{N}(p_L)$ belongs to one surface and the maximum, $l_{\max}$ to the other. If $l_{\max} - l_{\min} \leq 1$ holds then there are points only from one surface in $\mathcal{N}(p_L)$; otherwise, we need to determine which of the two surfaces p_L belongs to. First, the points are segmented into two groups. One of the two groups contain the points that satisfy the condition, $\left|l_{\max} - l_{q_L}\right| \leq 1$ and the other contains the points that satisfy the condition, $\left|l_{\min} - l_{q_L}\right| \leq 1$; here, q_L is a member of $\mathcal{N}(p_L)$. We find the average truncated color distance (defined later in this Section) of p_L to each group. p_L belongs to one of the two group that has color distance smaller than the other.

Occluded point with the highest priority has the smallest homogeneity estimate. We define homogeneity in the following way. p_L is homogeneous with its neighbor q_L if intensity of q_L is within 3*sigma of the mean intensity of p_L. Otherwise, the neighbor is non-homogeneous. Accordingly, homogeneity, $H(p_L, q_L)$ is defined by the following mixture model,

$$H(p_L, q_L) = \begin{cases} P\left(\left|\bar{I}(p_L) - I(q_L)\right|\right), \text{ if } \left|\bar{I}(p_L) - I(q_L)\right| \leq 3\mu_{LL}^{-1} \\ P\left(3\mu_{LL}^{-1}\right), \text{otherwise.} \end{cases} \tag{37}$$

Homogeneity of a site p_L with its neighborhood, $\mathcal{N}(p_L)$, $H(p_L, \mathcal{N}(p_L))$ is defined as

$$H(p_L, \mathcal{N}(p_L)) = \prod_{q_L \in \mathcal{N}(p_L)} H(p_L, q_L). \tag{38}$$

Since we are concerned with only a point with its relative (highest in this case) homogeneity, a negative log-homogeneity,

$$-\log H(p, \mathcal{N}(p)) = A + \sum_{q \in \mathcal{N}(p)} \psi(p, q), \tag{39}$$

is enough and also convenient for the estimation. In eq. (39), A is a constant due to normalization factor of $P(.)$. $\psi(p_L, q_L)$ is a cut-off function

$$\psi(p_L,q_L)=\begin{cases}\left|\bar{I}(p_L)-I(q_L)\right|, \text{ if } \left|\bar{I}(p_L)-I(q_{LL})\right|\le 3\mu_{LL}^{-1}\\ 3\mu_{LL}^{-1}, \text{otherwise.}\end{cases} \tag{40}$$

'*A*' does not affect a search for occluded point of the smallest homogeneity and therefore, can be ignored. So, $H(p_L,\mathcal{N}(p_L))$ can be instead replaced with $\sum_{q_L\in\mathcal{N}(p_L)}\psi(p_L,q_L)$. Accordingly, the highest homogeneity, $H(p_L,\mathcal{N}(p_L))$, corresponds to the smallest estimate $\sum_{q_L\in\mathcal{N}(p_L)}\psi(p_L,q_L)$.

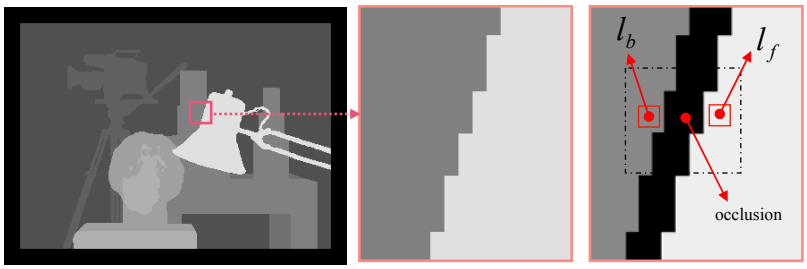
It was mentioned earlier that if $\mathcal{N}(p_L)$ has points from two surfaces, $\mathcal{N}(p_L)$ is divided into two neighborhoods (i.e., surfaces) $\mathcal{N}_1(p_L)$ and $\mathcal{N}_2(p_L)$ such that $\mathcal{N}(p_L)=\mathcal{N}_1(p_L)\cup\mathcal{N}_2(p_L)$. One of these two neighborhoods works as the set of control points in the interpolation of the disparity of the occluded points. We define the average color distance as

$$D(p_L,\mathcal{N}_i(p_L))=\frac{1}{\left|\mathcal{N}_i(p_L)\right|}\sum_{qL\in\mathcal{N}(pL)}\psi(p_L,q_L). \tag{41}$$

If $D(p_L,\mathcal{N}_1(p_L))<D(p_L,\mathcal{N}_2(p_L))$, then we pick $\mathcal{N}_1(p_L)$ as the set of control points; otherwise, we pick $\mathcal{N}_2(p_L)$. Once the control points are selected the parameters of the planar surface comprised of the disparities of the control points are estimated in the same way described before in WLS based algorithm; only exception is that the weights are assumed to be 1.0.

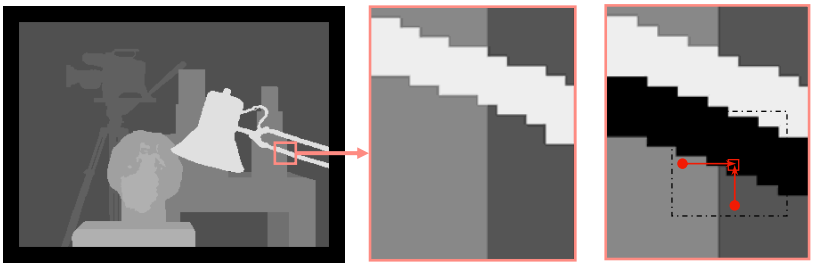
4.4.2 Summary of SLS algorithm

In this section, we summarize the proposed SLS occlusion filling algorithm with a flow diagram (Figure 4.14). In filling the occlusions, first, we find occluded point with the highest priority. Then we apply the constraint $\left|l_{pLb}-l_{qL}\right|\le 1\vee l_{qL}<l_{pLf}$ to find an initial set of control points. These control points form only one surface if the occlusions are created only by a background and foreground surfaces. We call the background surface, S_0, and take all its neighbors as control points. Otherwise, we assume that there are two background surfaces. The minimum and maximum disparity of the neighbors are taken as two seed points. We apply the disparity gradient constraint on these two seed points and extract the two background surfaces, S_1 and S_2. We measure color distance of these two surfaces from the occluded point. The surface with color distance smaller than the other one contains the occluded point. The neighbors that belong to this surface are the control points that participate in the interpolation.



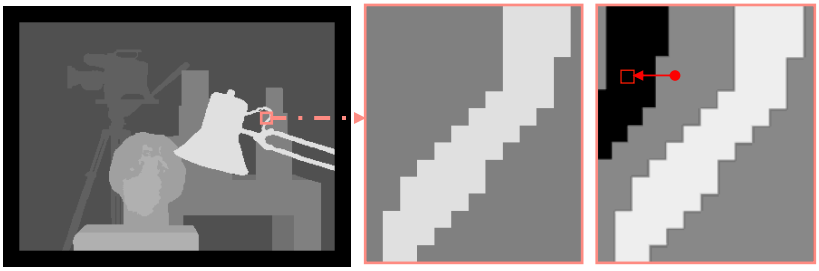
(a) (b) (c)

Figure 4.10: Occlusion with one background; (a) ground truth disparity of Tsukuba, (b) a zoomed-in portion of the disparity map, (c) neighborhood in occlusion created in the background by a foreground surface.



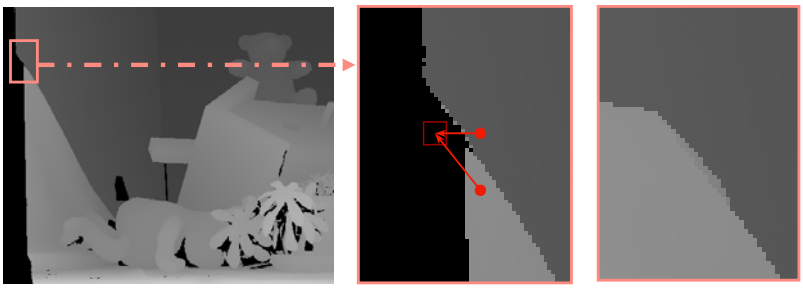
(a) (b) (c)

Figure 4.11: Occlusion when two background surfaces are present; (a) ground truth disparity map of the image Tsukuba, (b) a zoomed-in portion of the disparity map, (c) possibility of occluded point to be in one of the background surfaces.



(a) (b) (c)

Figure 4.12: Occlusions created by narrow objects; (a) ground truth disparity map of the image Tsukuba, (b) a zoomed-in portion of the disparity map, (c) occlusion created in the background by a narrow object.



(a) (b) (c)

Figure 4.13: Border occlusion filling needs a filling order; (a) ground truth disparity map of the image Tsukuba, (b) a zoomed-in portion of the disparity map, (c) ground truth disparity map.

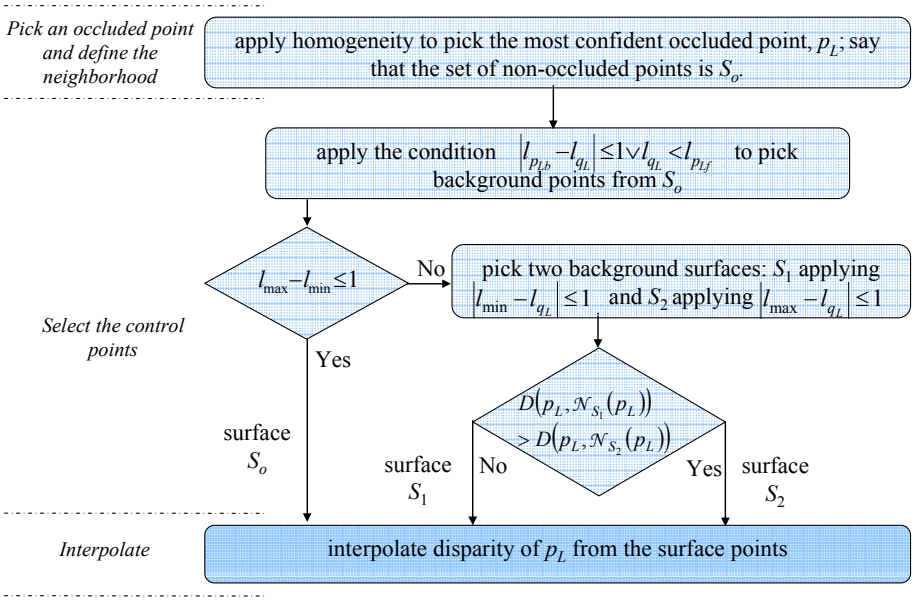


Figure 4.14: Detailed flow diagram of the proposed SLS occlusion filling algorithm

5 Experimental results

Experiments with the matching algorithm proposed in chapter 3 and occlusion filling algorithms proposed in chapter 4 are carried out on ground truth datasets prepared by the Middlebury College. We also perform experiments to observe computational efficiency of the proposed parameter estimation algorithms on BP, currently the best performer according to Middlebury evaluation. As mentioned earlier, many state of the art algorithms are tested against these datasets to study their performances. The comparative performances of these algorithms are published in the website, [Middlebury]. Currently there are four test pairs called Tsukuba, Venus, Teddy, and Cones for evaluation. An older version of the evaluation table includes Tsukuba, Venus, Sawtooth, and Map. All these images are shown in Figure 5.1. Radial distortions in these images are corrected and their ground truth disparity maps are obtained from laser scan. These images vary in degree of scene complexity, disparity range, color structure, etc. For instance, Tsukuba has the highest scene complexity among these four images – it has large un-textured to small texture-rich regions. Maximum disparities of all of these stereo pairs are known; they are listed in Table 5.1.

Figure 5.1: Six test images provided online by Middlebury College

Table 5.1: Maximum disparity ranges of the test images

Images	Tsukuba	Venus	Sawtooth	Map	Teddy	Cones
Maximum disparity (in pixels)	15	19	19	29	59	59

5.1 Stereo matching results

In this section, results generated by the proposed stereo matching algorithm are presented with performance study on the ground truth datasets. Figure 5.2 shows visual comparison of our results with results from Zhang-Seitz [Zhang07] and Cheng-Caelli [Cheng07] algorithms, the two algorithms other than ours' that also estimate the parameters statistically. This comparison is for the old dataset in the Middlebury College, which includes Tsukuba, Venus, Sawtooth, and Map. In Figure 5.3, we compare the results for two images Teddy and Cones which are included in the new Middlebury dataset in addition to Tsukuba and Venus. In this comparison, we do not include results of Chen-Caelli [Cheng07] algorithm, which was evaluated by the authors only for the old dataset. Numeric representations of these results are provided in Table 5.2 and Table 5.3. These tables show the percentages of points that match incorrectly. As mentioned earlier, a point is incorrectly matched if its absolute disparity difference with the ground truth is not more than one. A rainbow color coding of the ground truth disparity maps and disparity maps generated by the proposed matching algorithm is presented in Figure 5.4.

Two different statistical evaluations for matching error are done using the Middlebury stereo images. The first evaluation includes points only in the non-occluded regions, i.e. regions that are visible in both cameras. The second includes points only in the discontinuous regions. Discontinuous regions do not include the occluded points. Based on the two evaluations, the algorithms are ranked for comparisons. A rank is simply the chronological position after sorting the algorithms from the lowest to the highest according to their matching performance on one ground truth. In this way, an algorithm has eight rank numbers on four image pairs, each of them with two ground truths. For each algorithm, the rank numbers are added and averaged to find a 'score'. Then, all the algorithms are sorted again to find their comprehensive performance positions. Table 5.2 is for the test image pairs of the old dataset with performances evaluated according to the old evaluation table. Table 5.3 is for the new test image pairs in the new dataset with performances evaluated according to the new evaluation table. The two numbers in each cell under the columns with image names have two entries. The first one is the error rate and the second one is the rank. 'Score' of each algorithm is shown in the rightmost columns.

Until now the best performing local algorithm is proposed by Yoon and Kweon [Yoon06]. Their algorithm, however, is unable to perform well on images with repetitive patterns. For instance, in 'Map' pair their algorithm has 1.13% of matching error in non-occluded regions, whereas our proposed algorithm has matching error of 0.32%. Besides, the parameters are estimated automatically in our algorithm; whereas, in [Yoon06], the parameters are set manually.

5.2 Matching results for LC-SEM images

We applied our matching algorithm on LC-SEM stereo images. LC-SEM is a new and emerging class of Scanning Electron Microscope (SEM) with large chamber for conducting nondestructive tests. Meter-scale large objects can be scanned in micro- or nano-scale using LC-SEM; images delivered by the scanning process are grayscale. For scanning, a sample, which could be conducting or non-conducting, is placed inside the LC-SEM chamber through an airtight door. The air is pumped out and then, an electron gun emits a beam of high energy electrons; the beam travels through a series of magnetic lenses that focus the electrons to a very fine spot. Near above the sample surface, a set of scanning coils moves the focused electron beam back and forth across the surface of the sample, row by row. As the electron hits each spot on the sample, secondary electrons are knocked out of the surface. A detector counts these electrons and sends the signal to an amplifier. Pixel intensities in the final grayscale image are obtained from the number of electrons emitted from each spot on the sample. Note that stereo images from SEM or LC-SEM can be reconstructed only up to the duality, which means that the real 3D surface could be inverted.

Figure 5.6 shows a LC-SEM stereo image pair. The LC-SEM stereo image acquisition is performed in a controlled geometry. The tilt angle between the two view points and the physical dimensions of the pixels of the images are known. High magnification of the images allows the use of affine transformation. The principal points of both images are chosen to appear approximately at the centers of the images. Imaging in such a constrained and known set-up enables us to model the scene in the Euclidian space without calibration of the microscope. Before matching we rectify the images. Since the images are captured at high magnification (400X), the distortion effect on surface geometry due to the perspective projection is negligible. Therefore, we can assume the image formation as an affine process. Due to the affine assumption and the fact that the object is tilted only around the X-axis of the LC-SEM, the images become rectified after a clockwise 90^{o} rotation (Figure 5.6I and (d)). Figure 5.7(a) shows disparity map of the LC-SEM stereo image pair generated by the proposed algorithm. Figure 5.7(b) shows the rainbow color-coded disparity map. Details of the theory of the 3D reconstruction are presented in the appendix.

5.3 Runtime performance

Table **5.4** shows runtime (in minutes) of 144 iterations of the proposed stereo matching algorithm described in Chapter 3. The first row mentions the name of the stereo pair with its size and maximum disparity. It is evident from the table that while the runtime depends on the image size and maximum disparity, it also depends on the scene structure. Venus and Sawtooth have almost similar size and disparity range, but they direr in runtime significantly. Figure 5.8 shows graph plots of percentages of matching error in non-occluded regions and runtime versus number of iterations.

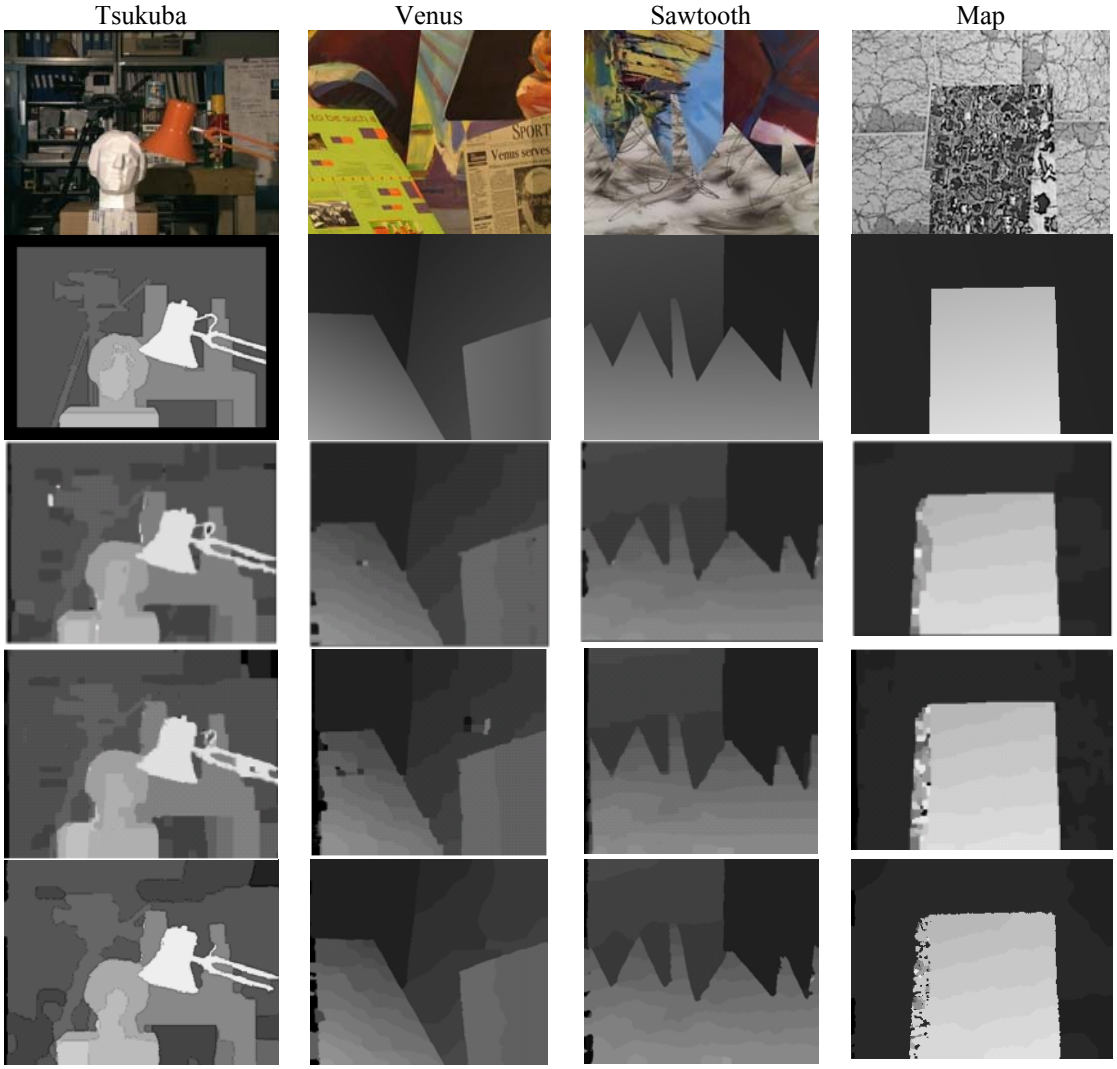


Figure 5.2: Disparity maps of the Middlebury test images Tsukuba, Venus, Sawtooth, and Map (from left to right; 1st row: test images, 2nd: ground truth disparity maps, 3rd: disparity maps of Zhang-Seitz

[Zhang07], 4th: Cheng-caelli [Cheng07], and 5th: proposed algorithm [Huq08b]

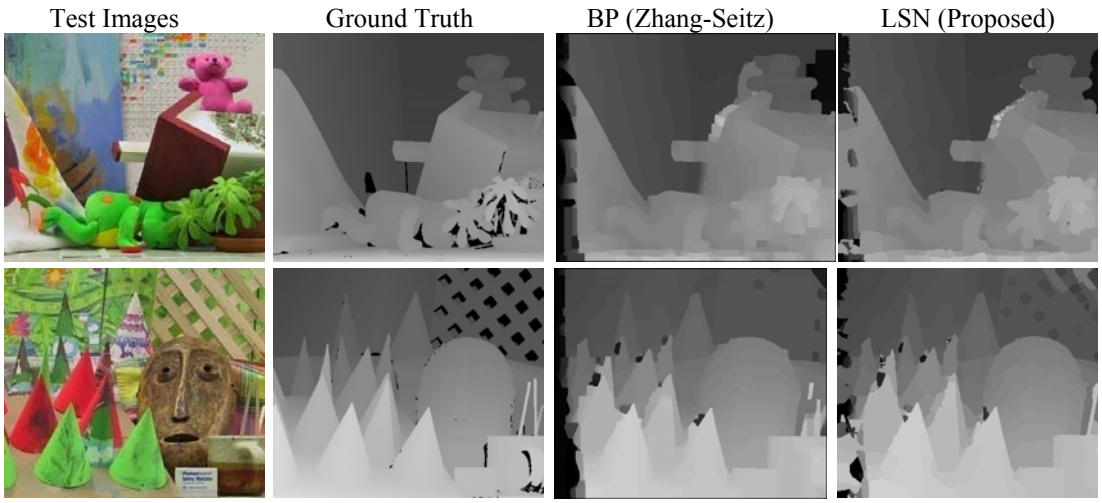
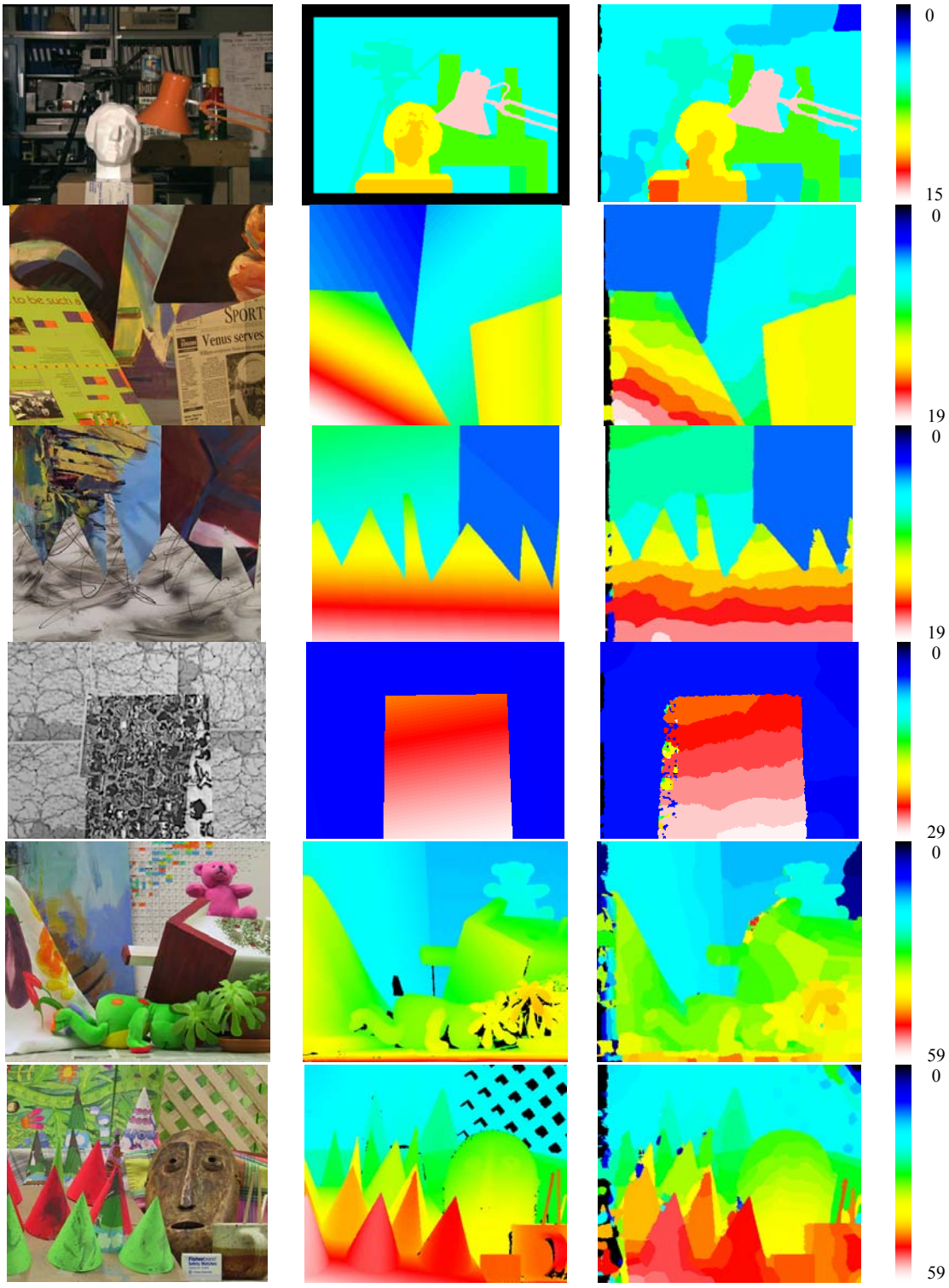


Figure 5.3: Disparity map of the Middlebury test images Teddy and Cones; 1st column: test images, 2nd: ground truth disparity map, 3rd: disparity maps of Zhang-Seitz [Zhang07], and 4th: proposed algorithm [Huq08b]



(a) (b) (c)

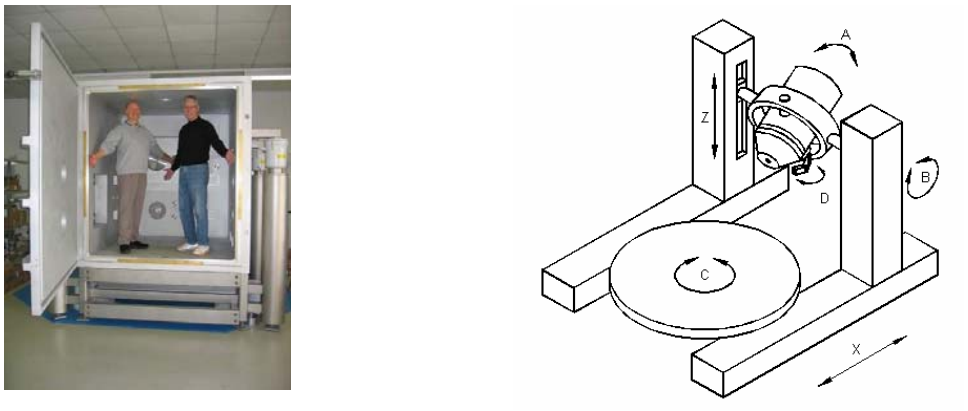
Figure 5.4: Disparity map results on the middle bury test dataset; (a) six test images Tsukuba, Venus, Sawtooth, Map, Teddy, and Cones (from top to bottom), (b) ground truth disparity map, (c) disparity map generated by our proposed algorithm.

Table 5.2: Comparison of percentage of matching error for non-occluded and discontinuous regions in the Middlebury old test image set – Tsukuba, Venus, Map, and Sawtooth

Algorithms	Error statistics regions	Tsukuba	Venus	Map	Sawtooth	Score
Proposed [Huq08b])	Non-occluded	1.53 13	0.97 10	0.34 13	0.32 7	9.87
	Discontinuous	5.38 3	6.05 14	4.81 15	3.20 5	
BP (Zhang-Seitz) [Zhang07]	Non-occluded	1.87 20	1.53 20	0.20 2	0.83 14	12.0
	Discontinuous	7.13 9	10.37 21	2.20 1	3.48 9	
BP (Cheng-Caelli) [Cheng07]	Non-occluded	3.65 28	3.41 32	0.10 1	1.23 21	21.1
	Discontinuous	15.33 32	17.40 29	1.33 1	7.91 25	

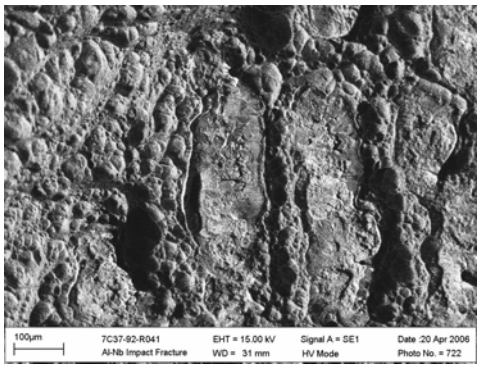
Table 5.3: Comparison of percentage of matching error (as of April 2008) for non-occluded and discontinuous regions in the Middlebury new test image set – Tsukuba, Venus, Teddy, and Cones

Algorithms	Error statistics regions	Tsukuba	Venus	Teddy	Cones	Score
Proposed [Huq08b]	Non-occluded	1.53 19	0.97 20	10.34 28	5.7 26	20.75
	Discontinuous	5.38 6	6.05 17	18.92 24	13.0 26	
BP (Zhang-Seitz) [Zhang07]	Non-occluded	1.87 22	1.53 26	12.27 30	6.02 26	25.5
	Discontinuous	7.13 18	10.37 25	22.79 29	15.08 28	

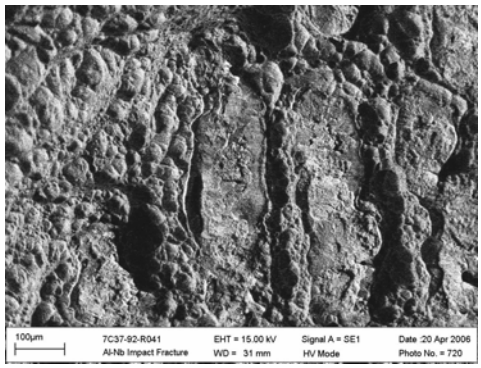


(a) (b)

Figure 5.5: (a) LC-SEM chamber and (b) LC-SEM diagram showing platform and electron gun with rotational and translational axes (images are provided by Y12, Oakridge National Research Laboratory)



(a)

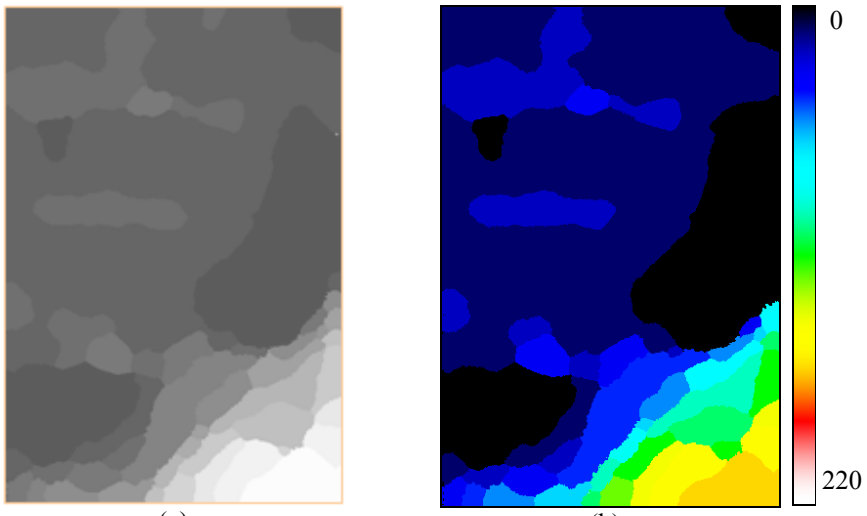


(b)

I

(d)

Figure 5.6: (a) and (b) are stereo image pair captured at 400X magnification with 6 degrees of tilt angle apart; image size is 1024×690 with each pixel 1µm×1µm in physical dimension; (c) and (d) are stereo image pair cut from images in (a), (b) and rectified with a clockwise 90^{o} rotation.



(a) (b)

Figure 5.7: Disparity map results on LC-SEM stereo images; (a) gray-coded and (b) color-coded disparity map generated by the proposed matching algorithm

Table 5.4: Runtime (in minutes) of 144 iterations of the proposed stereo matching algorithm

Tsukuba (384×288) 15	Venus (434×383) 19	Map (284×216) 29	Sawtooth (434×380) 19	Teddy (450×375) 59	Cones (450×375) 59
29	98	22	32	294	383

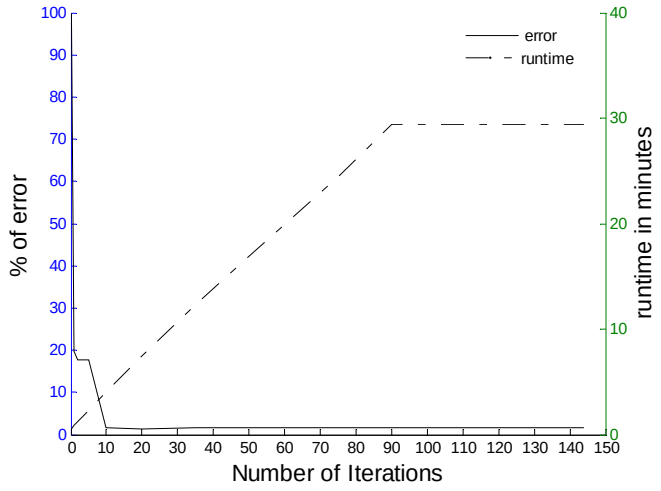


Figure 5.8: Graph plots for percentages of matching error and runtime vs. the number of iterations for stereo image pair Tsukuba.

5.4 Occlusion detection

Occlusions are the image regions that are visible in one of the stereo images but they are invisible in the other. Formally, these occlusions are called half occlusions, since they are visible in one and invisible in the other. There are algorithms that explicitly mark the occluded points during matching. A popular method is to match the images both ways and then mark those pixels as occlusions that do not have the same left-to-right and right-to-left labels. Since the proposed matching algorithm performs symmetric matching and make disparity maps from both ways readily available to us, the later technique is used to detect the occluded points. After matching is converged, we mark p_L in the left image as an occluded point if $l_{p_L} \neq -l_{p_R}$. Note that the signs of the labels are opposite.

Figure 5.9 shows three images, Tsukuba, Venus, and LC-SEM and their occlusions in the left image. Clearly, there are black points that are not along the borders but on the surface. These are the points that satisfy the condition, $l_{p_L} \neq -l_{p_R}$, since, the labels we work with are discrete.

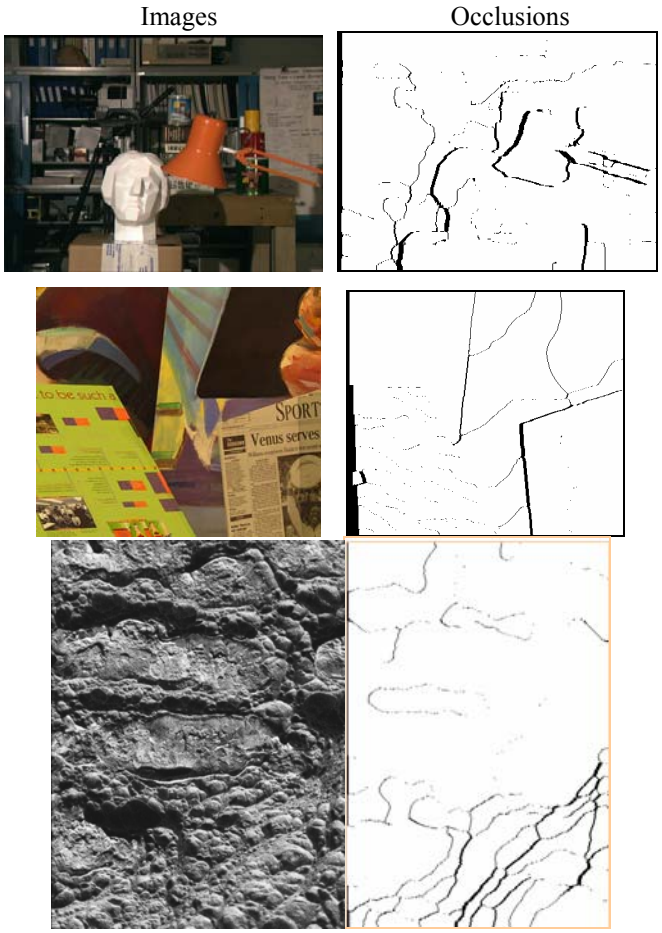


Figure 5.9: Top – Middlebury image 'Tsukuba' and its occlusions; middle – 'Venus' and its occlusions; bottom – LC-SEM image with occlusion

5.5 Occlusion filling results

Figure 5.10 shows the test image Teddy, its ground truth occlusion map, gray-coded disparity map, and color-coded disparity map. Figure 5.11 and Figure 5.12 show occlusion filling results in Teddy for four occlusion filling algorithms discussed in chapter 4. In Figure 5.11, the blue and red circles show the areas where NDA fills the occlusions inconsistently. In Figure 5.12(a), WLS do well in the mentioned red circular area but fails in occluded regions inside the blue circle. Disparity of the white region fails to spread consistently when using the WLS interpolation model. In Figure 5.12(b) and (c), the occlusion filled disparity maps generated by DIS and SLS are difficult to visually differentiate. In the next section we provide statistical evaluations of all four algorithms on all six test images.

5.6 Comparisons of the occlusion filling algorithms

We have described four algorithms NDA, DIS, WLS, and SLS with their visually presented performance on the test image, Teddy. As mentioned earlier, NDA is an existing method and DIS is inspired by one of the recent papers. The other two are developed by us. Table 5.5 lists statistical performance scores of these algorithms when applied on ground truth disparity maps of all six test images. The performance is evaluated as the percentage of error in occlusion filling, which appears as top entry in each cell of the table. To study the comparative performance, we sort the percentage of error rates from low to high and assign an algorithm its corresponding chronological position, which is written in the bottom of each cell of the table. For a particular algorithm, we average the chronological positions for all test cases and call it ‘score’. The lowest score indicates the best occlusion filling algorithm. The table shows that SLS is the best algorithm with score 1.33.

5.7 Occlusion filled disparity maps

We apply the segmentation based least squares (SLS) for interpolation to fill occlusions in the disparity maps generated by our proposed matching algorithm. Figure 5.12 shows results after occlusion filling. From left to the right, the figures show ground truth disparity maps, disparity maps of left-to-right and right-to-left matches, detected occlusion, and disparity maps after occlusion filling. For a better visualization of local disparity contrast, Figure 5.14 and Figure 5.15 show color-coded version of the disparity maps.

We numerically compare occlusion filled disparity maps of our proposed matching and occlusion filling algorithm (SLS) with disparity maps of existing algorithms listed in

the Middlebury performance table. Note that the main contribution of this dissertation is improved accuracy in matching among the existing algorithms that apply statistical estimation of the MRF parameters. The tables below rather emphasizes where an algorithm with automatic parameter estimation stands compared to others with manual settings of the parameters. Table 5.5 lists performancse of the proposed algorithm and a few other existing algorithms. The test dataset for this table, called Middlebury old test data, includes Tsukuba, Venus, Sawtooth, and Map. The table shows that the proposed algorithm performs better than average with the old dataset. On the other hand, Table 5.6 lists performance for the new dataset that includes Tsukuba, Venus, Cones, and Teddy. According to this table, the proposed algorithm performs almost average. Figure 5.16 and Figure 5.17 compare our results with the best algorithms respectively in old and new Middlebury evaluation tables.

5.8 Experiments with BP

The proposed parameter estimation techniques are implemented with belief propagation (BP) to study performance of the proposed parameter estimation technique in BP. We recall BP here with its equations for message update, data and smoothness term definitions, and assignments of the MRF parameters. Message in BP is defined and updated according to

$$m^{t}_{q_L p_L}\left(l_{p_L}\right) = \min_{l_{p_L}}\left(D_{q_L}\left(l_{q_L}\right) + \lambda V\left(l_{q_L}, l_{p_L}\right) + \sum_{s_L \in \mathcal{N}(q_L)\setminus p_L} m^{t-1}_{s_L p_L}\left(l_{q_L}\right)\right).$$

The data likelihood term is defined as $D_{p_L}\left(l_{p_L}\right) = \min(T_D, \left|I(p_L) - I(q_L)\right|)$ and the smoothness term as $V\left(l_{p_L}, l_{q_L}\right) = \min\left(T_{q_L}, \left|l_{p_L} - l_{q_L}\right|\right)$. A label l_{p_L} that minimizes the belief

$$b_{p_L}\left(l_{p_L}\right) = D_{p_L}\left(l_{p_L}\right) + \sum_{q_L \in \mathcal{N}(p_L)} m^{T}_{q_L p_L}\left(l_{p_L}\right), \tag{42}$$

is taken as the label for the pixel p_L. According to our estimation technique the MRF parameters λ_D, T_D, T_{q_L}, and λ_{q_L} are computed as

$\lambda_D = \mu_{LL}$, $T_D = 3\mu_{LL}^{-1}$, $T_{q_L} = \max\left(T_{\Delta I_{LR}}, 3\nu_{q_L}^{-1}\right)$ with $T_{\Delta I_{LR}} = 1.0$, and $\lambda_{q_L} = \nu_{q_L}$.

The MRF parameters are estimated and updated alternate with the matching iterations. In the existing algorithm, BP starts with a set of initial values for the parameters and finds the disparity map. The parameters are updated from the current disparity map using EM iterations and then applied again to refine the disparity map.
Since the EM iterations are nested inside the matching iteration, computational cost of the overall matching algorithm is high. The EM estimation iterates several times to update the parameters. From experiments, we found that approximately, 8 EM iterations are needed for convergence of the parameters as well as for the matching algorithm to deliver the best result in percentage of matching error. According to our proposed estimation, λ_D and T_D are estimated *a priori* from one of the stereo images. Besides, as described in this

dissertation, the smoothness model parameters are estimated using a combination of maximum likelihood and disparity gradient constraint.

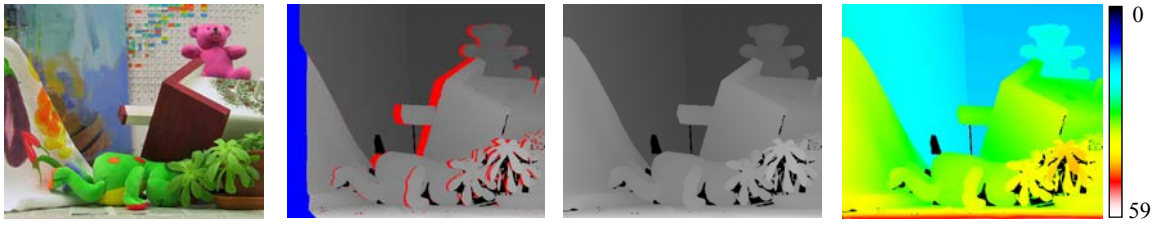
With BP, we implement an existing algorithm for fast message update, proposed by Falzenwalb and Huttenlocher [Falzenwalb04]. While standard BP estimates each smoothness term in an order of two of the number of disparity labels, without sacrificing any matching accuracy Falzenwalb and Huttenlocher estimated the smoothness terms with an order of one. We call BP with acceleration algorithm of Falzenwalb and Huttenlocher [Falzenwalb04] EMEM since both the data and smoothness parameters are estimated using the EM algorithms applying the parameter estimation of Zhang and Seitz [Zhang07]. Then, we conduct three experiments with each experiment providing some acceleration in computational performance. For the evaluation we use the new test dataset which includes Tsukuba, Venus, Teddy, and Cones. We call our experiments acceleration step 1, acceleration step 2, and acceleration step 3.

With the proposed parameter estimation implemented on top of EMEM, in acceleration step 1, we conduct experiment applying noise equivalence (NE) only in the parameter estimation of data likelihood; we also call this experiment NEEM. Table 5.8 lists performance results of EMEM and NEEM. For each of the experiment the table, there are three rows. The first row shows computational time of EMEM; the second row shows time gain of the NEEM experiments with respect to EMEM, i.e. the ratio of computational times of an experiment and EMEM; the third row shows matching error for non-occluded and discontinuous regions. In all experiments, BP is iterated 160 times. The lowest matching error rates for a particular test are shown in bold face. Considering the error rates for all the test images, it is obvious that maintaining comparable matching performance the computational time costs come down significantly.

In acceleration step 2, the proposed approach for the acceleration is different. Before presenting results of this experiment, we introduce the proposed approach in detail here. In BP, when all the messages of a site converge, the messages do not need to be further updated. Messages are converged if they remain unchanged for two or more consecutive iterations; this means that computational time can be saved by avoiding unnecessary message update. Yang et al. has applied message comparisons, i.e. checking if $\Delta m_t = \left| m^t_{qLpL}\left(l_{pL}\right) - m^{t-1}_{qLpL}\left(l_{pL}\right) \right|$ equals to zero, to determine convergence of the pixel p_L to save computational time in BP [Yang06b]. If Δm_t equals to zero in some iteration, p_L does not need further message update. However, the messages take some time to become mature when they do not change further. In our experiments, we skip initial message comparisons to avoid unnecessary checking for convergence and thus, improve the computational time further. Figure 5.19 shows results on some experiments with skipping message comparisons in initial iterations for all four images of the Middlebury new dataset. Each graph shows that as we skip the message comparisons in the initial iterations, the runtime for 160 iterations goes down initially; at the same time, the matching error also reduces in most of the experiments. We perform two experiments with NEEM algorithm: one is by skipping no iterations (NEEMS0) and the other is by skipping 40 initial iterations (NEEMS40) with the number of initial iterations, 40, chosen empirically from the graphs in Figure 5.19. Table 5.9 lists performance results of NEEMS0 and NEEMS40.

In acceleration step 3, we repeat the experiments on NEEMS40 with ML estimation of the smoothness parameters instead of EM as proposed in [Zhang07]. We call these ML based experiments NEMLS40. Table 4.6 lists performance results of NEEMS40 and NEMLS40. This experiments show that the percentage of error is increased due to replacement of the smoothness parameters by the ML estimation. However, the gain in computational time is significant. The time-accuracy compromise is justified for many real-time applications, for instance obstacle avoidance and tracking, where time is more important than gaining highly accurate solution.

For the test image 'Cones', Figure 5.20 shows plots of percentage of matching error along the left y axis and computation times along the right y axis at various numbers of BP iterations for all the above experiments. The plot shows that with comparable percentages of matching error rates throughout the graphs, proposed acceleration algorithms spend computational times less than EMEM.



(a) (b) (c) (d)

Figure 5.10: (a) Stereo image Teddy, (b) border occlusion (in blue) and non-border occlusion (in red) of Teddy, (c) gray-coded ground truth disparity map, and (d) color-coded ground truth disparity map

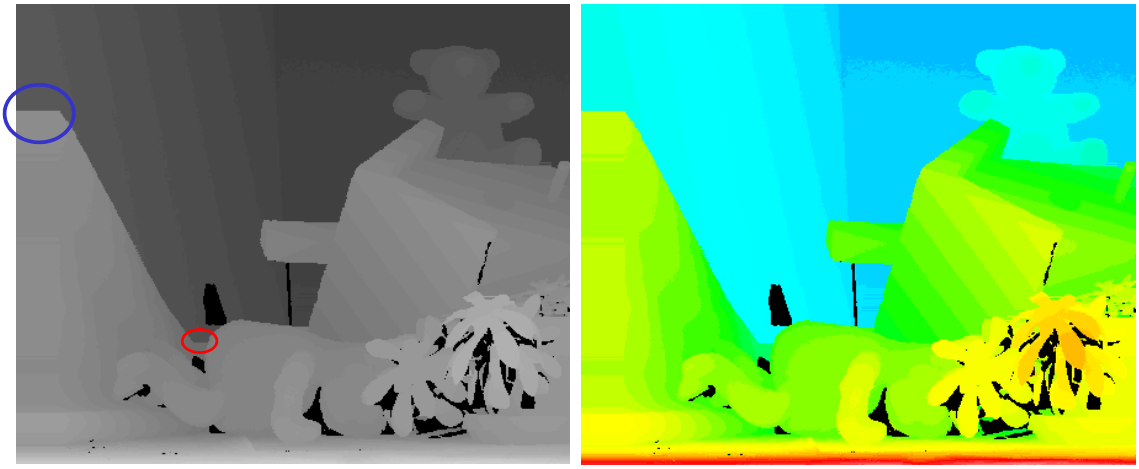


Figure 5.11: Gray- and color-coded disparity maps after filling occlusions in ground truth disparity maps with NDA

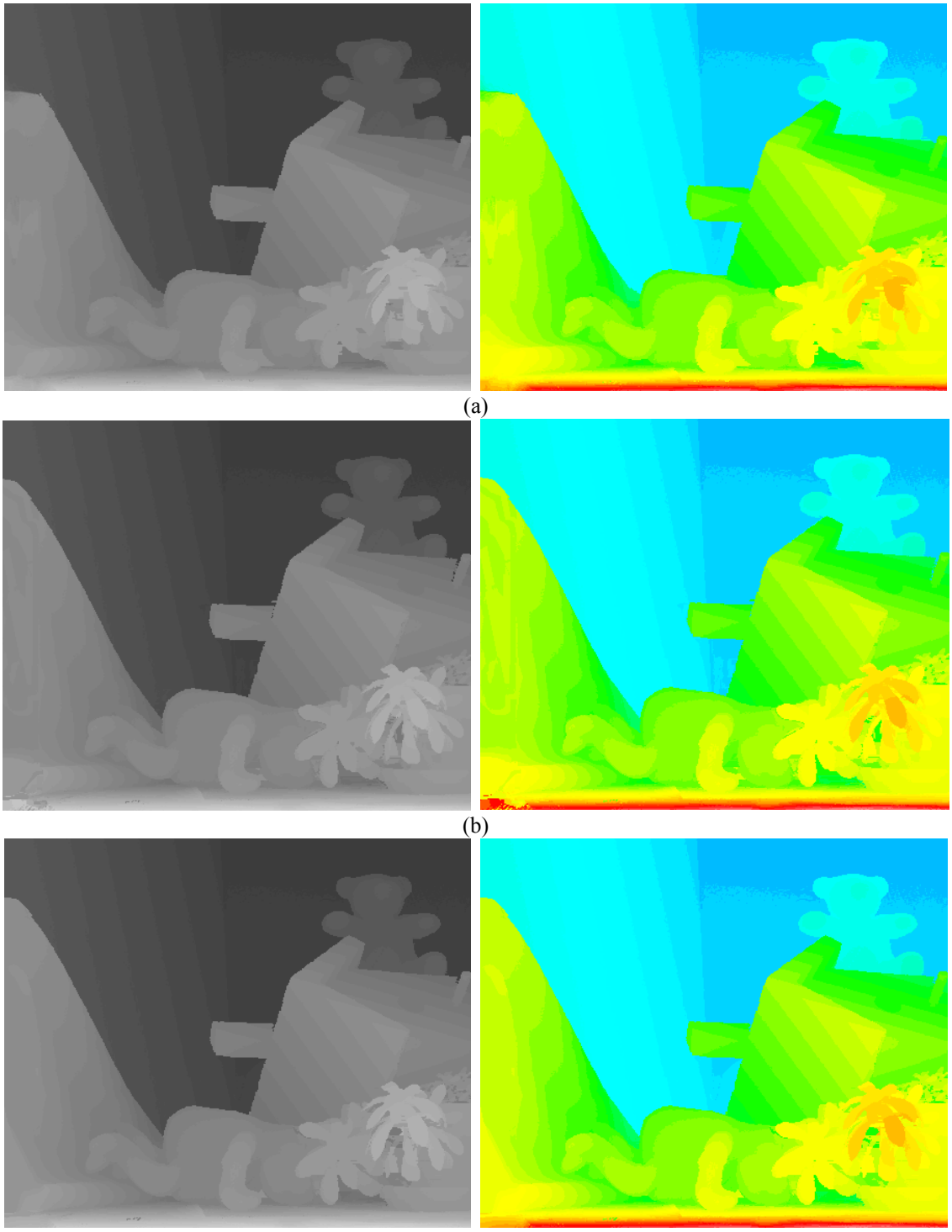


Figure 5.12: Gray- and color-coded disparity maps after filling occlusions in ground truth disparity maps with (a) WLS, (b) DIS, and (c) SLS

Table 5.5: Percentages of error in occlusion filling applying NDA, DIS, WLS, and SLS

Algorithms	Tsukuba	Venus	Cones	Teddy	Sawtooth	Map	Score
Neighbor's Disparity Assignment (NDA) [Yang06]	7.8 3	5.5 4	**32.1** 1	24.6 2	4.7 3	10.2 4	2.71
Diffusion in Intensity Space (DIS) [Min08]	**3.98** 1	4.66 2	39.29 4	27.2 4	2.72 2	1.75 2	2.5
Weighted Least Squares (WLS)	8.50 4	5.09 3	34.95 3	24.85 3	7.87 4	4.67 3	3.33
Segmentation based Least Squares (SLS)	5.49 2	**4.58** 1	34.89 2	**16.26** 1	**1.74** 1	**0.87** 1	**1.33**

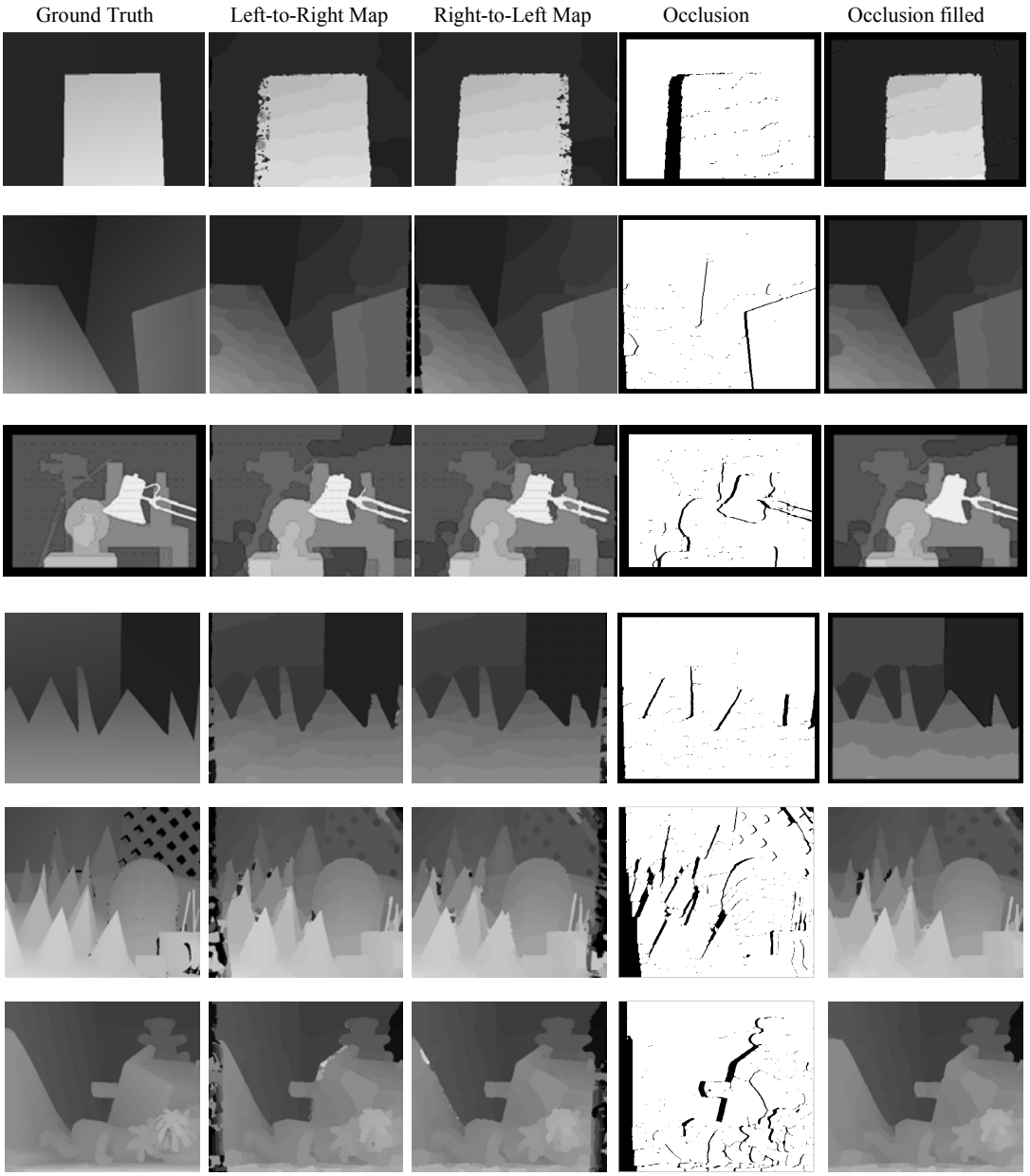


Figure 5.13: Disparity maps, occlusions, and occlusion filling results on the Middlebury College test images Map, Venus, Tsukuba, Sawtooth, Cones, and Teddy (from top to bottom). Occlusions are filled with SLS linear interpolation model.

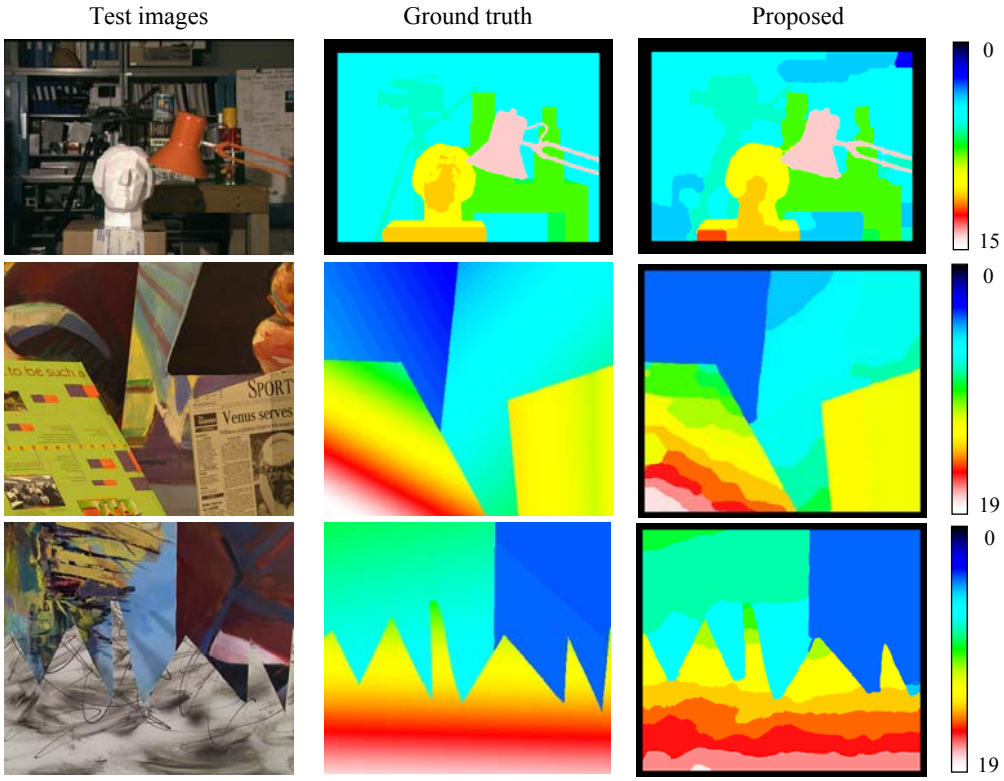


Figure 5.14: Color-coded disparity map on the middle bury test images Tsukuba, Venus, and Sawtooth; left: test images, middle: ground truth, and right: our results; the right most column shows disparity scale in color.

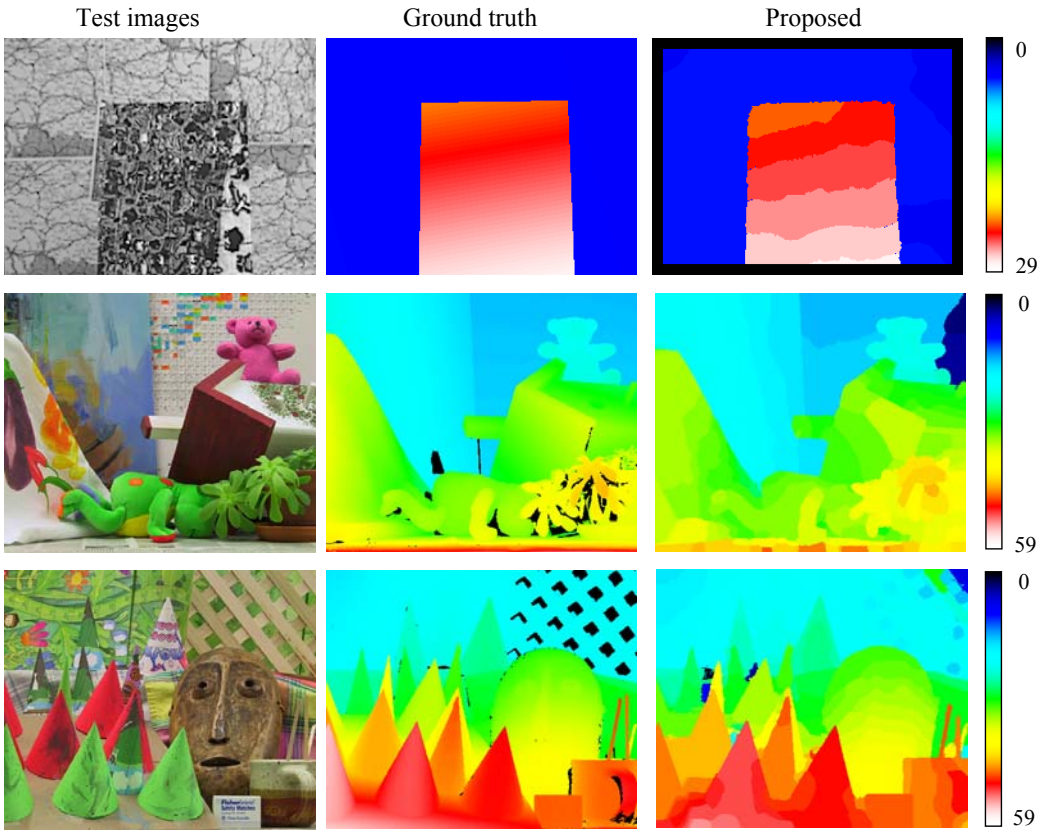


Figure 5.15: Color-coded disparity map on the middle bury test images Map, Teddy, and Cones; left: test images, middle: ground truth, and right: our results with color-coding scale.

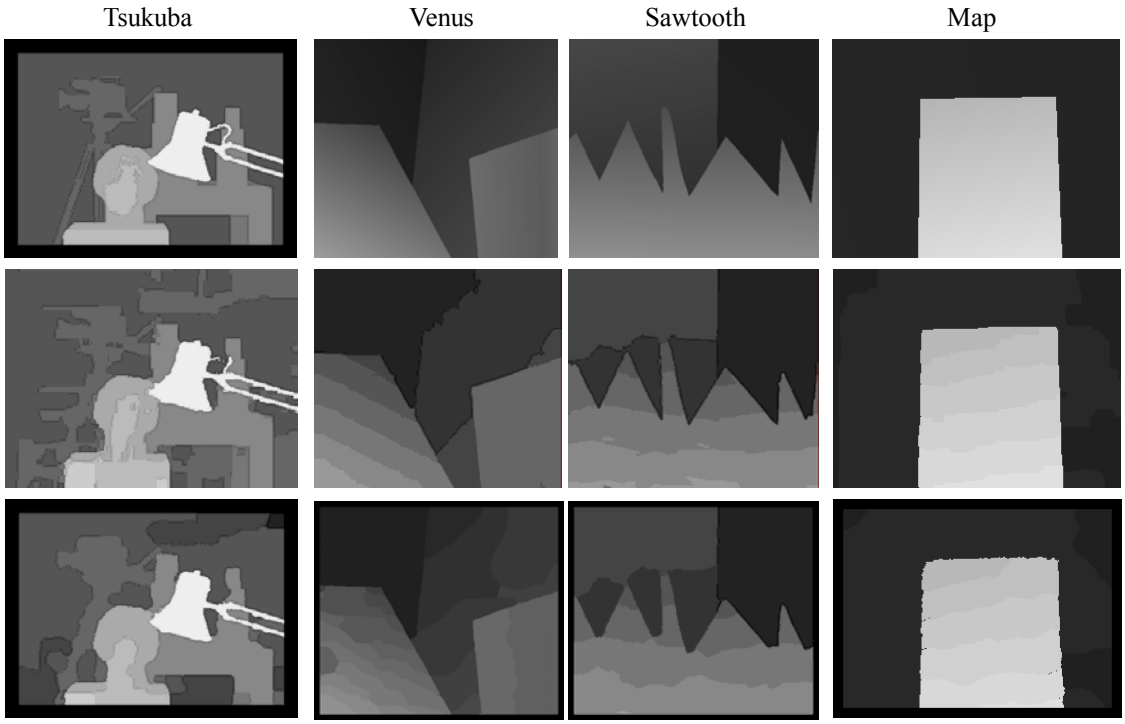


Figure 5.16: Disparity maps; top row: ground truth, middle: SymBP+Occ [Sun05], and bottom: proposed

Table 5.6: Percentages of error in disparities for non-occluded, discontinuous, and all regions in Tsukuba, Venus, Sawtooth, and Map (Middlebury old evaluation dataset), after filling occlusions applying SLS.

Algorithms	Tsukuba			Venus			Sawtooth			Map			Score
	non-occ	disc.	All	non-occ	disc.	All	non-occ	disc.	All	non-occ	disc.	All	
SymBP + Occ [Sun05]	0.97 2	5.45 2	- -	0.16 4	2.77 6	- -	0.19 1	2.09 1	- -	0.16 1	2.20 1	- -	2.75
Proposed	1.53 13	5.38 2	2.3 -	0.97 10	6.05 14	1.67 -	0.32 7	3.2 10	0.89 -	0.34 13	4.81 15	0.68 -	**10.5**
Segm.-based GC [Hong2004]	5.08 32	11.94 22	- -	9.44 39	8.2 30	- -	4.06 35	11.90 28	- -	1.84 32	10.22 25	- -	11
2 Pass DP [Kim2005]	1.53 13	8.25 14	- -	0.94 9	5.72 13	- -	0.61 10	5.25 14	- -	0.70 20	9.32 21	- -	13.85
SO [Christopher06]	5.08 32	11.94 22	- -	9.44 39	8.2 30	- -	4.06 35	11.90 28	- -	1.84 32	10.22 25	- -	30.37

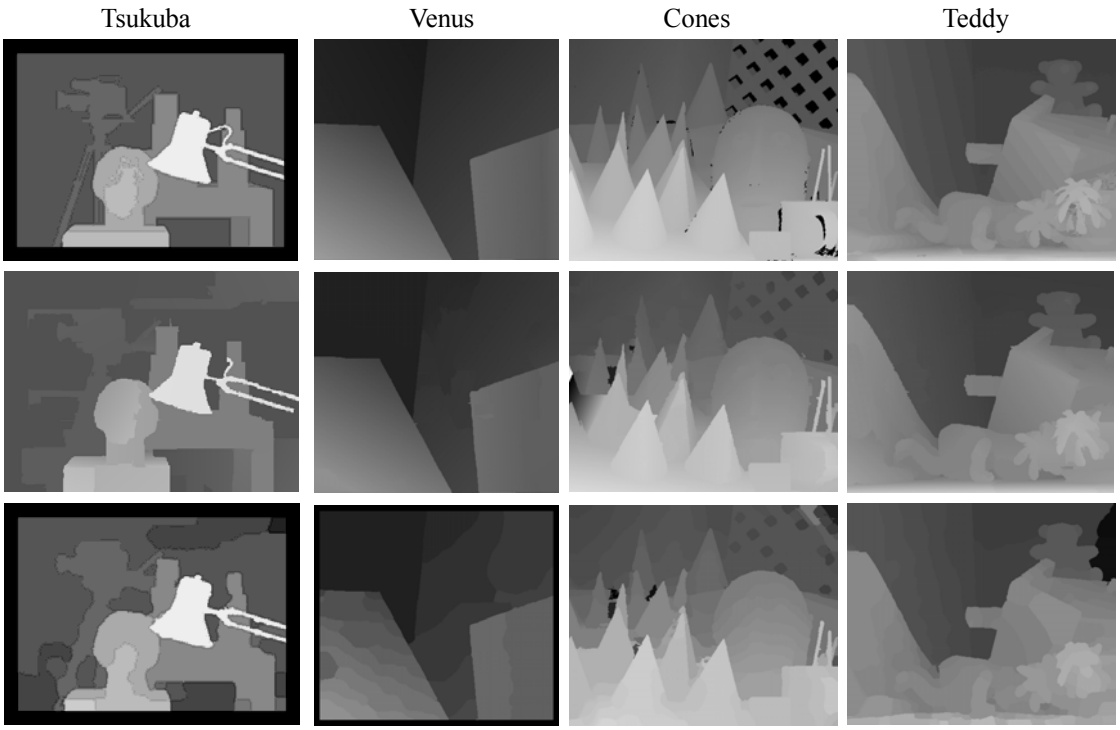


Figure 5.17: Disparity maps; top row: ground truth, middle: AdaptiveBP [Klaus06], and bottom: proposed

Table 5.7: Percentages of error in disparities for non-occluded, discontinuous, and all regions in Tsukuba, Venus, Cones, and Teddy (Middlebury new evaluation dataset), after filling occlusions applying SLS.

Alg.	Tsukuba			Venus			Cones			Teddy			Score
	non-occ	disc.	All	non-occ	disc.	All	non-occ	disc.	All	non-occ	disc.	All	
AdaptingBP [Klaus06]	1.11 7	5.79 8	1.37 3	0.10 1	1.44 1	0.21 3	2.48 1	7.92 2	7.32 3	4.22 3	11.8 3	7.06 2	3.1
RegionTreeDP [Lei06]	1.39 21	6.85 19	1.64 9	0.22 9	1.93 7	0.57 12	7. 42 34	16.8 24	11.9 28	6.31 18	11.8 14	11.9 12	17.3
Proposed	1.53 23	5.38 7	2.3 24	0.97 26	6.05 23	1.67 26	5.7 32	13.0 33	11.43 27	10.34 34	18.92 27	12.99 18	**25**
GC+occ [Kolmogorov01]	1.19 9	6.24 12	2.01 21	1.64 32	6.75 25	2.19 31	11.2 29	19.8 33	17.4 29	5.36 26	13.0 27	12.4 28	26.8
SO [Christopher06]	5.08 44	12.2 37	7.22 45	9.44 46	21.9 42	10.9 46	13.0 45	22.3 43	22.8 46	19.9 46	26.3 41	28.2 47	44

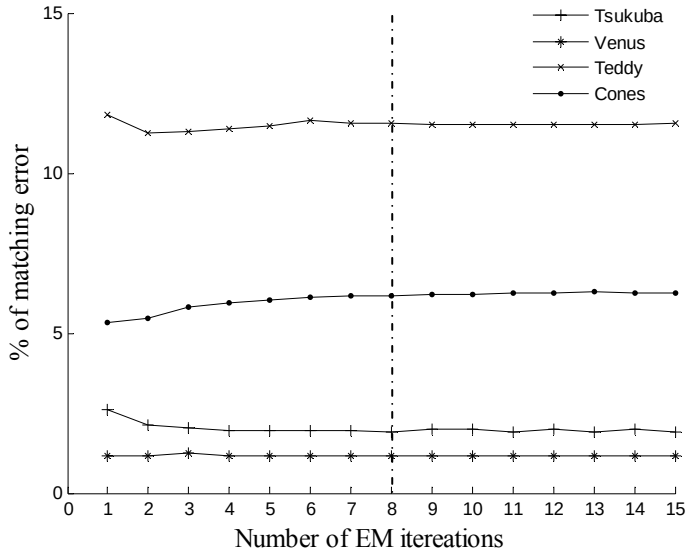


Figure 5.18: Graphs for percentages of matching error vs. number of EM iterations for four ground truth stereo pairs

Table 5.8: Matching and computational time performance of existing and proposed parameter estimation algorithms (number of BP iterations = 60)

Algorithms	Performance criteria	Tsukuba	Venus	Teddy	Cones
EMEM (existing) Zhang-Seitz [Zhang07, Felzenswalb04].	Time (in seconds)	202	342	747	740
	Time Gain: EMEM/EMEM	1.0	1.0	1.0	1.0
	Error (%): Non-occ, discontinuity	**1.92**, **9.59**	**1.18**, **15.55**	11.55, 24.58	6.15, 16.13
NEEM (proposed) Data parameter replaced with noise equivalence (NE) [Zhang07, Felzenswalb04, Huq08c]	Time (in seconds)	133	226	546	545
	Time Gain: NEEM/EMEM	1.5	1.5	1.4	1.4
	Error (%): Non-occ, discontinuity	2.12, 10.49	1.34, 17.63	**11.16**, **23.46**	**5.91, 15.47**

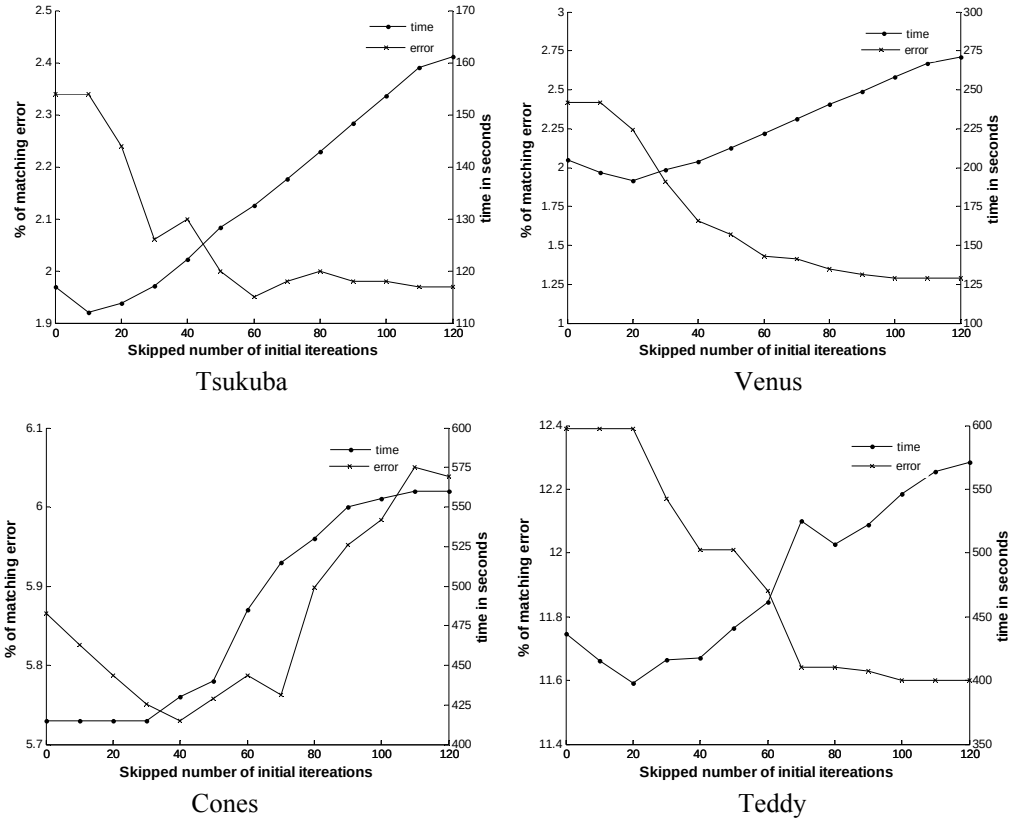


Figure 5.19: Percentages of matching error and computational time in seconds versus skipped number of initial iterations with message comparisons.

Table 5.9: Matching and computational time performances of existing and proposed parameter estimation algorithms with message comparison and message comparison skipping (number of BP iterations = 60)

Algorithms	Performance criteria	Tsukuba	Venus	Teddy	Cones
NEEMS0 Message comparison [Zhang07, Felzenswalb04, Yang06]	Time (in seconds)	117	205	437	483
	Time Gain: NEEMS0/EMEM	1.7	1.7	1.7	1.5
	Error (%): Non-occ, discontinuity	2.24, **9.07**	2.42, **17.66**	12.39, 23.70	5.73, 15.07
NEEMS40 (proposed) Message comparison after initial 40 iterations [Zhang07, Felzenswalb04, Yang06, Huq08c]	Time (in seconds)	74	126	316	311
	Time Gain: NEEMS40/EMEM	2.7	2.7	2.4	2.4
	Error (%): Non-occ, discontinuity	**2.00**, 9.63	**1.81**, 18.18	**12.19**, **23.45**	**5.58**, **14.51**

Table 5.10: Matching and computational time performances of the proposed algorithms with and without applying proposed estimation of the smoothness parameters

Algorithms	Performance criteria	Tsukuba	Venus	Teddy	Cones
NEEMS40 (proposed) Message comparison after initial 40 iterations [Zhang07, Felzenswalb04, Yang06, Huq08c]	Time (in seconds)	74	126	316	311
	Time Gain: NEEMS40/EMEM	2.7	2.7	2.4	2.4
	Error (%): Non-occ, discontinuity	**2.00, 9.63**	1.81, 18.18	**12.19, 23.45**	**5.58, 14.51**
NEMLS40 (proposed) ML estimation of the smoothness parameters [Zhang07, Felzenswalb04, Yang06, Huq08c]	Time (in seconds)	57	103	296	254
	Time Gain: NEMLS40/EMEM	3.5	3.3	2.5	2.9
	Error (%): Non-occ, discontinuity	3.48, 11.04	**1.80, 17.12**	13.42, 24.22	5.83, 14.52

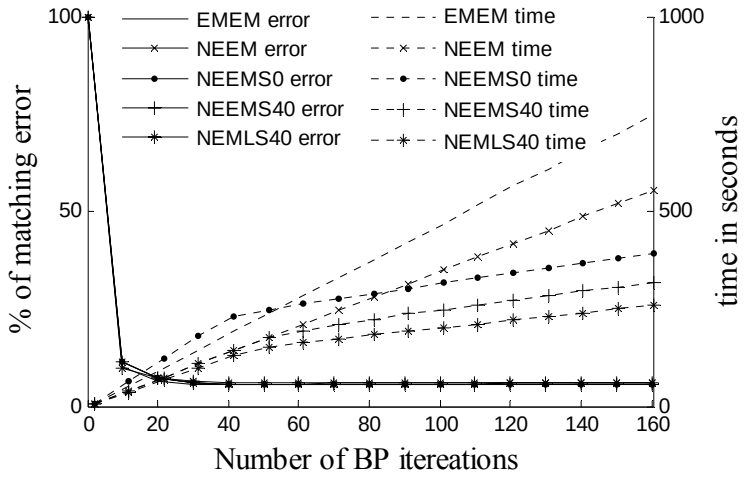


Figure 5.20: Percentages of non-occluded matching error and computational time versus number of BP iterations for 'Cones' for up to 160 iterations

6 Conclusions and future work

In this dissertation, we proposed a novel local optimization algorithm for dense MRF stereo matching. Using a concise cost function, the energy parameters are defined explicitly and estimated statistically. Symmetric cost functions are developed and matching is performed symmetrically for improved accuracy and also to detect occlusions. We introduce a concept called "strictness of neighborhood" in order to balance between the data and smoothness costs and apply the statistical parameters directly into the cost function. The data likelihood parameters are estimated from one of the images applying noise equivalence hypothesis that the distributions of the *within image* and *between image* noises are equivalent. Our approach estimates the data likelihood parameters *a priori* to avoid the costly nested iteration and to enforce convergence in matching by eliminating interdependency between the data and smoothness parameters. Two basic smoothness model parameters are estimated using a combination of maximum likelihood and disparity gradient constraint. Data and smoothness discontinuity (i.e. occlusions) handling parameters are also statistical and dependent on their respective data and smoothness model parameters. An adaptive support neighborhood technique is incorporated for global convergence of the algorithm. At the same time, the adaptive support neighborhood provides additional capability in discontinuity handling by including enough homogeneous points in the neighborhood. Smoothness model parameters of the neighbors in the support neighborhood are obtained from the two basic smoothness model parameters applying observed linear relationship among the parameters. A pair of symmetric cost functions for symmetric matching is developed. Symmetric matching reduces matching error and at the same time allows detection of the occlusions. The proposed algorithm is applied to ground truth test datasets provided online by the Middlebury College and comparable results are obtained. Results show that our algorithm performs better than existing algorithms that apply statistical estimation of the parameters. The proposed algorithm performs well on LC-SEM stereo images which are noisier than images captured by the regular cameras.

We proposed several algorithms for occlusion filling; we conducted experiments with existing and proposed algorithms on ground truth disparity maps of the Middlebury test images and compared their performance. Experiments show that our best proposed algorithm, which is segmentation based least squares with linear model of extrapolation, performs better than the existing algorithms. The proposed occlusion filling algorithm is independent to the matching algorithm, hence, can be applied in occlusion filling in disparity maps delivered by any stereo matching algorithm.

Finally, we implement our parameter estimation algorithm for belief propagation to study performance of the estimation algorithm. Applying noise equivalence only for likelihood parameter estimation, we obtain results comparable with existing algorithms', yet computational complexity is greatly reduced. Additionally, by applying maximum likelihood for the prior model parameters we gain in time complexity further.

While our focus is on an effective modeling and automatic estimation of MRF parameters and matching for offline matching applications, there are real time stereo matching algorithms in the literature [Gong05, Wang06, Forstmann04] that could benefit from our proposed algorithm. Since parameter estimation is usually time consuming, to save runtime the real time algorithms use hard-coded parameters determined empirically; hence, they are not robust with changes in outdoor scene. In this dissertation, we show that good matching results can be achieved by estimating the data likelihood parameters statistically and *a priori* from one of the images; thus hard coding of the parameters can be avoided to cope with changing scene. The neighborhood selection and symmetric matching techniques can be applied with existing well known algorithms such as graph cuts and belief propagation for potential increase in matching accuracy and other performance.

The existing algorithms as well as our proposed one estimate and apply the MRF parameters globally. The global estimation is valid for data model parameters. However, the smoothness parameters may not be consistent locally. A better approach would be local estimation of the parameters. A local estimation, however, is not as unbiased as the global one, mainly because the size of sample participating in the estimation is small. There are still works needed to be done in this direction. In Figure 6.1, it is evident that the image region of rectangle A is more distorted than the region in rectangle B when they are seen in the right image. The smoothness parameters should be determined independently for these two regions to obtain optimal results.

Figure 6.1: Left and right stereo images of Teddy and its ground truth disparity map

Bibliography

[Amidan05] B. G. Amidan, T. A. Ferryman, and S. K. Cooley, "Data outlier detection using the Chebyshev theorem," *IEEE Aerospace Conference*, pp. 3814–3819, Mar 2005.

[Audirac05] H. Audirac, A. Beloiarov, F. Núñez, and J. Villegas, "Dense disparity map based on STICA algorithm," *Expo Forestal*, Mexico, 2005.

[Baker81] H. Baker and T. Binford, "Depth from Edge and Intensity Based Stereo," *Int'l Joint Conf. on Artificial Intelligence*, Vancouver, Canada, pp. 631–636, 1981.

[Barnard89] S. Barnard, "Stochastic Stereo Matching over Scale," *Int'l Journal of Computer Vision* (IJCV), Vol. 3, No. 1, pp. 17–32, May 1989.

[Benshair96] A. Benshair, P. Miche, and R. Debrie, "Fast and automatic stereo vision matching algorithm based on dynamic programming method," *Pattern Recognition Letters*, Vol. 17, pp. 457–466, 1996.

[Besag86] J. Besag, "On the statistical analysis of dirty pictures (with discussion)," *Journal of the Royal Statistical Society*, Series B 48, pp. 259–302, 1986.

[Blake87] A. Blake and A. Zisserman. Visual Reconstruction, MIT Press, 1987.

[Bleyer04] M. Bleyer and M. Gelautz, "A layered stereo algorithm using image segmentation and global visibility constraints," *IEEE Int'l Conf. on Image Processing* (ICIP), Vol. 5, pp. 2997–3000, Oct 24-27, 2004.

[Boykov01] Y. Boykov, O. Veksler, and R. Zabih, "Fast Approximate Energy Minimization via Graph Cuts*," IEEE Transactions on Pattern Analysis and Machine Intelligence* (PAMI), Vol. 23, No. 11, pp. 1222–1239, 2001.

[Brockers05] R. Brockers, M. Hund, and B. Mertsching, "Stereo vision using cost-relaxation with 3D support regions," *Image and Vision Computing New Zealand* (IVCNZ), 2005.

[Burt80] P. Burt and B. Julesz, "Modifications of the classical notion of Panum's fusional limit," *Perception*, Vol. 9, pp. 671–682, 1980.

[Chen01] Q. Chen and G. Medioni, "Building 3–D Human Face Models from Two Photographs," *The Journal of VLSI Signal Processing, Kluwer Academic Publishers*, Vol. 27, No. 1–2, pp. 127–140, 2001.

[Cheng07] L. Cheng and T. Caelli, "Bayesian stereo matching," *Int'l Journal of Computer Vision and Image Understanding*, Vol. 106, pp. 85–96, 2007.

[Christopher06] Z. Christopher, S. Mario, and K. Konrad, "Scanline Optimization for Stereo on Graphics Hardware," *3rd Int'l Symposium on 3D Data Processing, Visualization and Transmission* (3DPVT), 2006.

[Comaniciu02] D. Comaniciu and P. Meer, "Mean shift: A robust approach towards feature space analysis," *IEEE Trans. on Pattern Analysis Machine Intelligence* (PAMI), Vol. 24, No. 5, pp. 603–619, 2002.

[Dempster77] A. Dempster, N. Laird, and D. Rubin, "Maximum likelihood from incomplete data via the EM algorithm ", *Journal of Royal Statistical Society*, Series B, Vol. 39, pp. 1–38, 1977.

[Deng06] Y. Deng and X. Lin, "A fast line segment based dense stereo algorithm using tree dynamic programming," *European Conf. on Computer Vision* (ECCV), Vol. 3, pp. 201–212, 2006.

[Elias56] P. Elias, A. Feinstein, and C. E. Shanon, "Note on maximum flow through a network," *IRE Transactions on Information Theory IT-2*, pp. 117–119, 1956.

[Etriby06] S. El-Etriby, A. Al-Hamadi, and B. Michaelis, "Dense depth map reconstruction by phase difference-based algorithm under influence of perspective distortion," *Special Issue of the International Journal Machine Graphics and Vision* (ICCVG), 2006.

[Etriby07] S. El-Etriby, A. Al-Hamadi, and B. Michaelis, "Dense stereo correspondence with slanted surface using phase-based algorithm," *IEEE Int'l Symposium on Industrial Electronics* (ISIE), Jun 2007.

[Faugeras93] O. Faugeras. Three Dimensional Computer Vision. MIT Press, 1993

[Felzenswalb04] P. Felzenswalb and D. Huttenlocher, "Efficient Belief Propagation for Early Vision," *Proc. of the IEEE Computer Society Conf. on Computer Vision and Pattern Recognition* (CVPR), Vol. 1, pp. 261-268, 2004.

[Forstmann04] S. Forstmann, J. Ohya, Y. Kanou, A. Schmitt, and S. Theuring, "Real-time stereo by using dynamic programming," *CVPR 2004 Workshop on real-time 3D sensors and their use*.

[Fusiello 98] G. A. Fusiello, E. Trucco, and A. Verri, "A compact algorithm for rectification of stereo pairs," *Machine Vision and Applications*, Vol. 12, No. 1, pp. 16–22, Jul 2000.

[Ford56] L. R. Ford and D. R. Fulkerson, "Maximum flow through a network," *Canadian Journal of Mathematics*, Vol. 8, pp. 399–404, 1956.

[Geiger91] D. Geiger and F. Girosi, "Parallel and Determinstic Algorithms from MRFs: Surface Reconstruction," *IEEE Trans. Pattern Analysis Machine Intelligence* (PAMI), Vol. 13, pp. 401–412, 1991.

[Gelatt83] S. Kirkpatrick, C. D. Gelatt, M. P. Vecchi, "Optimization by Simulated Annealing," *Science*, Vol. 220, No. 4598, pp. 671–680, 1983.

[Gelman03] A. Gelman, J. B. Carlin, H. S. Stern, D. B. Rubin. Bayesian Data Analysis. Second Edition, Publishers: Chapman and Hall, 2003.

[Geman84] D. Geman and S. Geman, "Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images,". *IEEE Trans. Pattern Analysis Machine Intelligence* (PAMI), Vol. 6, pp. 721–741, Nov 1984.

[Gong01] M. Gong and Y. Yang, "Multi-resolution stereo matching using genetic algorithm," *IEEE Workshop on Stereo and Multi-Baseline Vision*, 2001.

[Gong05] M. Gong and Y.-H. Yang, "Near real-time reliable stereo matching using programmable graphics hardware," *Proc. of the IEEE Computer Society Conf. on Computer Vision and Pattern Recognition* (CVPR), Vol. 1, pp. 924–931, 2005.

[Hao07] W. Hao, S. Huq, D. Page, B. Abidi, A. Koschan, and M. Abidi, "Nano-Scale 3D Metrology for Surface Characterization and Inspection of High-Precision Manufactured Components," *ANS/ENS International Meeting*, Washington, DC, Nov 11–15, 2007.

[Hartley04] R. Hartley and A. Zisserman. Multiple View Geometry in Computer Vision. Second Edition, Cambridge University Press, March 2004.

[Hemmleb97] M. Hemmleb and M. Schubert, "Digital Microphotogrammetry – Determination of the Topography of Microstructures by Scanning ElectronMicroscope," *Second Turkish- German Joint Geodetic Days*, Berlin, Germany. Pp. 745-752, May 1997.

[Hirschmüller05] H. Hirschmüller, "Accurate and efficient stereo processing by semi-global matching and mutual information," *IEEE Computer Society Conf. on Computer Vision and Pattern Recognition* (CVPR), Vol. 2, pp. 807–814, 2005.

[Hirschmüller06] H. Hirschmüller, "Stereo vision in structured environments by consistent semi-global matching," *Proc. of the IEEE Computer Society Conf. on Computer Vision and Pattern Recognitio*n (CVPR), Vol. 2, pp. 2386–2393, 2006.

[Holland75] J. H. Holland. Adaptation Natural and Artificial Systems, Ann Arbor, Michigan Univ., Michigan Press, 1975.

[Hong04] L. Hong and G. Chen, "Segment-based stereo matching using graph cuts," *Proc. of IEEE Computer Society Conf. on Computer Vision and Pattern Recognition* (CVPR), Vol. 1, pp. I-74–I-81, 2004.

[Huq04] S. Huq, B. Abidi, A. Goshtasby, and M. Abidi, "Stereo Matching with Energy Minimizing Snake Grid for 3D Face Modeling," *Proc. of SPIE, Defense and Security Symposium*, 2004, Florida, USA, Vol. 5404, pp.339–350, 2004.

[Huq06] S. Huq, B. Abidi, C. Kammerud, M. Abidi, J. Frafjord, and S. Deckanich, "3D Measurements of Wear on Machining Tools Using a Confocal Microscope," *Int'l Conf. on Microscopy and Microanalysis jointly with Int'l Metallographic Society* (IMS), Vol. 34, No. 2, Aug 2006.

[Huq07a] S. Huq, B. Abidi, S. Kong, and M. Abidi. Chapter 2: A Survey on 3D Modeling of Human Faces for Face Recognition. 3D Imaging for Safety and Security, pp. 25–67. Publisher: *Springer*, Jul 2007; ISBN: 978-1-4020-6181-3

[Huq07b] S. Huq, B. Abidi, D. Page, and M. Abidi, J. Frafjord, and S. Deckanich, "Inspection of Fracture Surfaces using 3D from Stereo Images of Large Chamber SEM*," Int'l Conf. on Microscopy and Microanalysis*, Florida, Aug 5–9, 2007.

[Huq07c] S. Huq, B. Abidi, and M. Abidi, "Stereo-based 3D Face Modeling using Annealing in Local Energy Minimization," *IEEE 14th Int'l Conf. on Image Analysis and Processing* (ICIAP), Modena, Italy, 10-13 Sep 2007.

[Huq08a] S. Huq, B. Abidi, D. Page, M. Abidi, J. Frafjord, and S. Deckanich, "3D Modeling from LC-SEM Stereo Images for Microscopic Surface Metrology and Material Characterization," *3rd Int'l Conf. on Design and Technology of Integrated Systems*, Mar 25–28, 2008.

[Huq08b] S. Huq, A. Koschan, B. Abidi, and M. Abidi, "MRF Stereo with Statistical Estimation of Parameters," *4th Int'l Symposium on 3D Data Processing, Visualization, and Transmission* (3DPVT), Atlanta, Jun 2008.

[Huq08c] S. Huq, A. Koschan, B. Abidi, and M. Abidi, "Efficient BP Stereo with Automatic Paramter Estimation," to appear in *Proc. of IEEE 15th Int'l Conf. on Image Processing* (ICIP), Oct 2008.

[Intlille94] S. S. Intlille and A. F. Bobick, "Disparity-Space Images and Large Occlusion Stereo," *Proc. of the 3rd European Conf. on Computer Vision* (ECCV), Vol. 2, 1994.

[Kammerud05] C. Kammerud, B. Abidi, S. Huq, and M. Abidi, "3D Nanovision for the Inspection of Micro-Electro-Mechanical Systems," *IEEE Int'l Conf. on Electronics, Circuits, and Systems*, Gammarth, Tunisia, Dec 2005.

[Kanade94] T. Kanade and M. Okutomi, "A Stereo Matching Algorithm with an Adaptive Window: Theory and Experiment," *IEEE Transactions on Pattern Analysis and Machine Intelligence* (PAMI), Vol. 16, No. 9, pp. 920–932, Sep 1994.

[Kim05] J. Kim, K. M. Lee, B. Choi, and S. Lee, "A dense stereo matching using two-pass dynamic programming with generalized ground control points," *Proc. of the IEEE Computer Society Conf. on Computer Vision and Pattern Recognition* (CVPR), Vol. 2, pp. 1075–1082, Jun 2005.

[Kittler05] J. Kittler, A. Hilton, M. Hamouz, and J. Illingworth, "3D Assisted Face Recognition: A Survey of 3D Imaging, Modelling, and Recognition Approaches," *Proc. of the IEEE Computer Society Conf. on CVPR*, Vol. 3, pp. 114–120, Jun 2005.

[Klaus06] A. Klaus, M. Sormann, and K. Karner, "Segment-based stereo matching using belief propagation and a self-adapting dissimilarity measure," *18th Int'l Conf. on Pattern Recognition*, Vol. 3, pp. 15–18, Aug 20–24, 2006.

[Kolmogorov01] V. Kolmogorov and R. Zabih, "Computing visual correspondence with occlusions using graph cuts," *IEEE Int'l Conf. on Computer Vision* (ICCV), Vol. 11, pp. 508–515, 2001.

[Kolmogorov02] V. Kolmogorov and R. Zabih, "Multi-camera scene reconstruction via graph cuts," *7th European Conf. on Computer Vision* (ECCV), Vol. 3, pp. 82–96, 2002.

[Kolmogorov06] V. Kolmogorov, "Convergent Tree-Reweighted Message Passing for Energy Minimization," *IEEE Transactions on Pattern Analysis and Machine Intelligence* (PAMI), Vol. 28, No. 10, pp. 1568–1583, Oct 2006.

[Lambert60] J. H. Lambert, "Photometria sive de mensure de gratibus luminis, colorum umbrae," Eberhard Klett, 1760.

[Larsen07] E. S. Larsen, P. Mordohai, M. Pollefeys, and H. Fuchs, "Temporally consistent reconstruction from multiple video streams using enhanced belief propagation," *11th IEEE Int'l Conf. on Computer Vision* (ICCV), Brazil, Oct 2007.

[Lauritzen88] S. L. Lauritzen and D. J. Spiegelhalter, "Local computations with probabilities on graphical structures and their application to expert systems," *Journal of Royal Statistical Society*, Series B, Vol. 50, pp. 157-224, 1988.

[Lei06] C. Lei, J. Selzer, and Y. Yang, "Region-tree based stereo using dynamic programming optimization," *Proc. of the IEEE Computer Society Conf. on Computer Vision and Pattern Recognition* (CVPR), pp. 2378–2385, Jun 2006.

[Li95] S. Z. Li. Markov Random Field Modeling in Computer Vision. Springer-Verlag, New York, 1995.

[Mattoccia07] S. Mattoccia, F. Tombari, and L. D. Stefano, "Stereo vision enabling precise border localization within a scanline optimization framework," *8th Asian Conference on Computer Visio* (ACCV), pp. 517–527, 2007.

[Meltzer05] T. Meltzer, C. Yanover, and Y. Weiss, "Globally Optimal Solutions for Energy Minimization in Stereo Vision Using Reweighted Belief Propagation," *Int'l Conf. on Computer Vision* (ICCV), pp. 428–435, 2005.

[Menger27] K. Menger, "Zur allgemeinen Kurventhoerie," *Fund. Math*, Vol. 10, pp. 96–115, 1927.

[Middlebury] http://vision.Middlebury.edu/~schar/stereo/newEval/php/results.php

[Midoh05] Y. Midoh, K. Miura, K. Nakamae, and H. Fujioka, "Statistical optimization of Canny edge detector for measurement of fine line patterns in SEM image," *Measurement Science and Technology*, Vol. 16, No. 2, pp. 477–487, 2005.

[Min08] D. Min and K. Sohn, "Cost Aggregation and Occlusion Handling With WLS in Stereo Matching," *IEEE Transactions on Image Processing*, Vol. 17, No. 8, pp. 1431-1442, Aug 2008.

[Mordohai06] P. Mordohai and G. Medioni, "Stereo using monocular cues within the tensor voting framework," *IEEE Trans. on Pattern Analysis and Machine Intelligence* (PAMI), Vol. 28, No. 6, pp. 968–982, 2006.

[Oho05] E. Oho, N. Baba, M. Katoh, T. Nagatani, M. Osumi, K. Amako, and K. Kanaya, "Application of the Laplacian filter to high-resolution enhancement of SEM images," *Journal of Electron Microscopy Technique*, Vol. 1, No. 4 , pp. 331–340, Feb 2005.

[Ohta85] Y. Ohta and T. Kanade, "Stereo by Intra- and Inter-scanline Search Using Dynamic Programming," *IEEE Trans. on Pattern Analysis and Machine Intelligence* (PAMI), Vol. 7, No. 2, pp. 139–154, 1985.

[Olague06] G. Olague, F. Fernández, C. Pérez, and E. Lutton, "The infection algorithm: an artificial epidemic approach for dense stereo correspondence," *Artificial Life*, Vol. 6, No. 4, pp. 593–615, 2006.

[Onofrio04] D. Onofrio, A. Sarti, and S. Tubaro, "Area matching based on belief propagation with applications to face modeling," *Int'l Conf. on Image Processing* (ICIP), pp. 1943–1946, 2004.

[Onofrio05] D. Onofrio, S. Tubaro, A. Rama, and F. Tarres, "3D Face Reconstruction with A Four Camera Acquisition System," *Int'l Workshop on Very Low Bit-Rate Video Coding*, Sardinia, Italy, 2005.

[Oriot98] H. Oriot and G. L. Besnerais, "Matching Aerial Stereo Images Using Graduated Non Convexity Techniques," *19th ISPRS Congress and Exhibition*, Jun 1998.

[Pearl88] J. Pearl. Probabilistic reasoning in intelligent systems: networks of plausible inference. *Morgan Kaufmann Publishers*, *Inc*., 1988.

[Peterson89] C. Peterson and B. Soderberg, "A new method for mapping optimization problems onto neural networks," *Int'l Journal of Neural Systems*, Vol. 1, No. 1, pp. 3–22, 1989.

[Pollard85] S. Pollard, J. Mayhew, and J. Frisby, "PMF: a stereo correspondence algorithm using a disparity gradient limit. *Perception*, Vol. 14, pp. 449–470, 1985.

[Potts52] R. B. Potts, "Some Generalized Order-Disorder Transformations," *Proc. of the Cambridge Philosophical Society*, Vol. 48, pp. 106−109, 1952.

[Rosenfeld76] A. Rosenfeld, R. Hummel, and S. Zucker, "Scene labeling by relaxation operation," *IEEE Transactions on Systems, Man, and Cybernetics*, Vol 6, pp. 267–287, 1976.

[Saito95] H. Saito and M. Mori, "Application of genetic algorithms to stereo matching of images," *Pattern Recognition Letters*, Vol. 16, pp. 815–821, 1995.

[Scharstein98] D. Scharstein and R. Szeliski, "Stereo Matching with Nonlinear Diffusion," *Int'l Journal of Computer Vision* (IJCV), Vol. 28, No. 2, pp. 155–174, 1998.

[Scharstein02] D. Scharstein and R. Szeliski, "A taxonomy and evaluation of dense two-frame stereo correspondence algorithms," *Int'l Journal of Computer Vision* (IJCV), Vol. 47, pp. 7–42, 2002.

[Serfling80] R. J. Serfling. Approximation Theorems of Mathematical Statistics. Wiley Series in Probability and Statistics, First Edition, 1980.

[Serfling02] R. J. Serfling. Approximation Theorems of Mathematical Statistics. Wiley Series in Probability and Statistics, Second Edition, 2001.

[Sun03] J. Sun, Na. Zheng, and H. Shum, "Stereo matching using belief propagation," *IEEE Trans. on Pattern Analysis and Machine Intelligence* (PAMI), Vol. 25, No. 7, pp. 787–800, Jul 2003.

[Sun05] J. Sun, Y. Li, S.B. Kang, and H.-Y. Shum, "Symmetric stereo matching for occlusion handling," *IEEE Conf. on Computer Vision and Pattern Recognition* (CVPR), Vol. 2, pp. 399–406, 2005.

[Szeliski06] R. Szeliski, R. Zabih, D. Scharstein, O. Veksler, V. Kolmogorov, A. Agarwala, M. Tappen, and C. Rother, "A Comparative Study of Energy Minimization Methods for Markov Random Fields," *9th European Conf. on Computer Vision* (ECCV), Vol. 2, pp. 19–26, Graz, Austria, May 2006.

[Tien04] F. Tien and C. Tsai, "Scanline-based stereo matching by genetic algorithms," *Int'l Journal of Production Research*, Vol. 42, No. 6, pp. 1083–1106, 2004.

[Tombari07] F. Tombari, S. Mattoccia, and L. Di Stefano, "Segmentation-based adaptive support for accurate stereo correspondence," *IEEE Pacific-Rim Symposium on Image and Video Technology* (PSIVT), 2007.

[Trucco98] E. Trucco and A. Verri. Introductory Techniques for 3D Computer Vision, Prentice–Hall, New Jersey, 1998.

[Tyler73] C. Tyler, "Stereoscopic vision: cortical limitations and a disparity scaling effect," *Science*, Vol. 181, pp. 276–278, 1973.

[Veksler05] O. Veksler, "Stereo correspondence by dynamic programming on a tree," *IEEE Conf. on Computer Vision and Pattern Recognition* (CVPR), Vol. 2, pp. 384–390, 2005.

[Wainwright05] M. J. Wainwright, T. S. Jaakkola, and A. S. Willsky, "MAP Estimation via Agreement on (Hyper)Trees: Message-Passing and Linear-Programming Approaches," *IEEE Trans. Information Theory*, Vol. 51, No. 11, pp. 3697–3717, Nov 2005.

[Wang06] L. Wang, M. Liao, M. Gong, R. Yang, and D. Nistér, "High-quality real-time stereo using adaptive cost aggregation and dynamic programming," *3rd Int'l Symposium on 3D Data Processing, Visualization and Transmission* (3DPVT), pp. 798–805, 2006.

[Wu83] C. F. J. Wu, "On the convergence properties of the EM algorithm," *Annals of Statistics*, Vol. 11, pp. 95–103, 1983.

[Yang06a] J. Y. Jang, K. M. Lee, and S. U. Lee, "Stereo matching using iterated graph cuts and mean shift filtering," *7th Asian Conf. on Computer Vision* (ACCV), Hyderabad, India, pp. 31–40, Jan 2006.

[Yang06b] Q. Yang, L. Wang, R. Yang, S. Wang, M. Liao, and D. Nistér, "Real-time global stereo matching using hierarchical belief propagation," *British Machine Vision Conference* (BMVC), pp. 989–998, 2006.

[Yang06c] Q. Yang, L. Wang, R. Yang, H. Stewénius, and D. Nistér, "Stereo matching with color-weighted correlation, hierarchical belief propagation and occlusion handling," *Proc. of the IEEE Computer Society Conf. on Computer Vision and Pattern Recognition* (CVPR), Vol. 2, pp. 2347–2354, 2006.

[Yang07] Q. Yang, R. Yang, J. Davis, and D. Nistér, "Spatial-depth super resolution for range images," *Proc. of the IEEE Computer Society Conf. on Computer Vision and Pattern Recognition* (CVPR), pp. 1–8, 2007.

[Yoon06] K.-J. Yoon and I.-S. Kweon, "Adaptive support-weight approach for correspondence search," *IEEE Trans. on Pattern Analysis and Machine Intelligence* (PAMI), Vol. 28, No. 4, pp. 650–656, 2006.

[Yoon07] K.-J. Yoon and I.-S. Kweon, "Stereo matching with the distinctive similarity measure," *11th IEEE Int'l Conf. on Computer Vision* (ICCV), Brazil, Oct 2007.

[Zhang05] L. Zhang and S. M. Seitz, "Parameter Estimation for MRF Stereo," *Proc. of IEEE Computer Society Conf. on Computer Vision and Pattern Recognition* (CVPR), San Diego, CA, Vol. 2, pp. 288–295, Jun 2005.

[Zhang07] L. Zhang and S. M. Seitz, "Estimating Optimal Parameters for MRF Stereo from a Single Image Pair," *IEEE Transactions on Pattern Analysis and Machine Intelligence* (PAMI), Vol. 29, No. 2, pp. 331–342, Feb 2007.

[Zitnick04] L. Zitnick, S. B. Kang, M. Uyttendaele, S. Winder, and R. Szeliski, "High-quality video view interpolation using a layered representation," *ACM Trans. on Graphics*, Vol. 23, No. 3, pp. 603–308, 2004.

[Zitnick07] L. Zitnick and S. B. Kang, "Stereo for image-based rendering using image over-segmentation," *Int'l Journal of Computer Vision* (IJCV), Vol. 75, No. 1, pp. 49–65, Oct 2007.

Appendix

Calibration of SEM/LC-SEM stereo imaging system

SEM/LC-SEM can provide stereo images of an object captured from two view points of known angle between them. In comparison with usual stereo system that uses more than one camera, in SEM/LC-SEM stereo imaging there is only one camera (equivalently the scanner). In SEM, the object is placed on a stage and then posed at different tilt angles by rotating the stage. In LC-SEM, both the stage and the scanner (equivalently the camera) can be repositioned. For reconstruction of the object from stereo images, the scanner can be modeled with affine projection if the objects are scanned at high magnification (300x or more). Due to affine projection assumption the intrinsic parameter, focal length, appears linearly proportional to the magnification. Assuming the magnification as 1.0 allows the reconstruction to happen up to a scale factor. The aspect ratio is assumed to be 1.0; hence, magnification remains the same along both the X and Y axes. Skew angle is 90 degrees. The extrinsic parameters of the system are the orientation and location of the scanner. In our setup, the orientation is known and the parameters related to the location are unknown. In our modeling, there are four parameters indirectly related to the location of the scanner. These parameters are determined from calibration.

To present the theory of calibration, we describe the SEM/LC-SEM imaging scenario with the following coordinate systems. We assume two coordinate systems – one attached to the scanner and the other to the stage. We also assume that the stage is located right under the scanner, i.e., principal axis of the scanner, which is also the Z-axis of the scanner coordinate system, goes through the origin of the stage. Z-axis of the stage is parallel with principal axis of the scanner. XY planes of the scanner and the stage are also parallel. Say, difference between the origins of the stage and scanner coordinate systems is T_z.

The coordinate systems for the scanner, stage, and the object are introduced in Figure A1. The object coordinate system has an origin located at a feature point of the object that can be easily detected in all stereo images (2 or more). We call this feature point an anchor point. XY planes of all three coordinate systems and their X- and Y-axes are parallel. The Z-axes are vertical and parallel with the principal axis of the scanner. With respect to the object coordinate system, let us say that the origin of the scanner coordinate system is (X_c, Y_c, Z_c) and origin of the stage coordinate system is (X_s, Y_s, Z_s). The object could be reconstructed with respect to the stage coordinate system. Instead, we choose the object coordinate system for two reasons. First, we want to self calibrate the system. Second, we want to gain higher accuracy in reconstruction by reconstructing the object in its own scale of dimension (micro or nano). With respect to the stage or the scanner coordinate system the object dimension is small which makes the numerical roundup or truncation error in depths relatively big.

Say at tilt 0, the object point (X, Y, Z) is affine projected at (x, y). Before the projection is applied, the object point is transformed into the scanner coordinate system only by a translation (X_c, Y_c, Z_c).

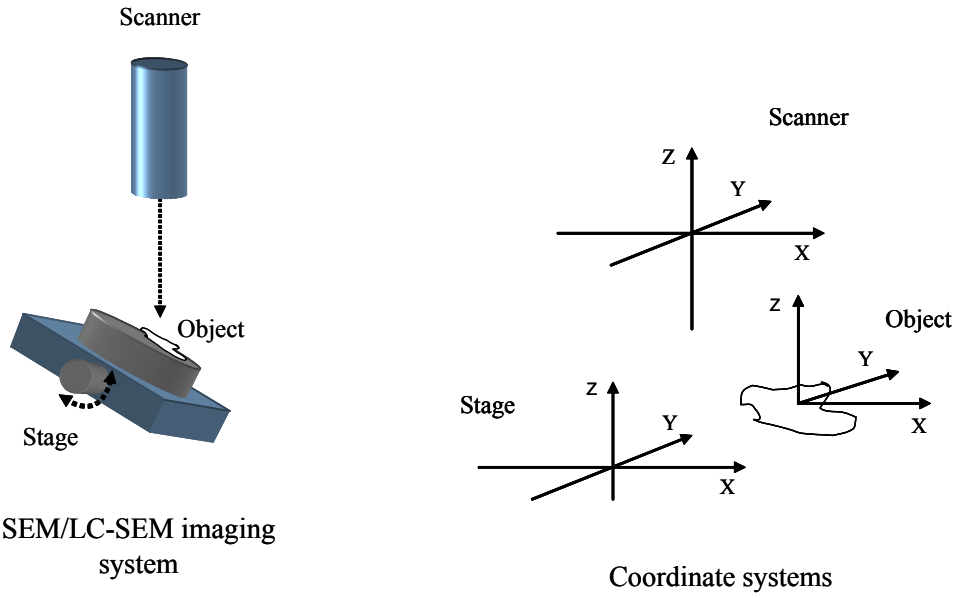


Figure A1: (a) SEM/LC-SEM stereo imaging set up and (b) scanner, stage, and object coordinate systems.

$$\begin{bmatrix} x \\ y \\ 0 \end{bmatrix} = \begin{bmatrix} M & 0 & 0 \\ 0 & M & 0 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} X - X_c \\ Y - Y_c \\ Z - Z_c \end{bmatrix}. \tag{43}$$

Say, at tilt β, the object point (X, Y, Z) is affine projected at (x', y'). Before the projection is applied, the object point is rotated by β around Y-axis. Usually, such rotation moves the object out of the field of view of the scanner. A translation (T_x, 0, 0) is applied to bring the scene back into the view. The rotation changes the relative coordinates of the origins of the stage and the scanner coordinate systems, which are with respect to the object coordinate system. The new scanner location after rotation of the object is $\begin{bmatrix} X_c - X_S - T_x \\ Y_c - Y_S \\ Z_c - Z_S \end{bmatrix} + R_Y(\beta) \begin{bmatrix} X_S \\ Y_S \\ Z_S \end{bmatrix}$, which is obtained by first transforming all the coordinates into stage coordinate system, then applying rotation β and translation (T_x, 0, 0), and then transforming all the coordinates back into the object coordinate system. Thus, we find the following projection relation for the second view.

$$\begin{bmatrix} x' \\ y' \\ 0 \end{bmatrix} = \begin{bmatrix} M & 0 & 0 \\ 0 & M & 0 \\ 0 & 0 & 0 \end{bmatrix} \left(R_Y(\beta) \begin{bmatrix} X \\ Y \\ Z \end{bmatrix} - \begin{bmatrix} X_c - X_S - T_x \\ Y_c - Y_S \\ Z_c - Z_S \end{bmatrix} - R_Y(\beta) \begin{bmatrix} X_S \\ Y_S \\ Z_S \end{bmatrix} \right). \tag{44}$$

From (43) and (44), we have

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ \cos\beta - 1 & 0 & -\sin\beta \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \end{bmatrix} = \begin{bmatrix} X_c + x \\ Y_c + y \\ -\sin\beta . Z_s + (\cos\beta - 1) . X_s - T_x + x' - x \end{bmatrix}, \tag{45}$$

where, M = 1.0 is assumed letting the reconstruction to happen up to a scale factor. In (45), there are 4 unknowns X_c, Y_c, Z_s, and X_s, which can be estimated from correspondence of one point in 3 views or one and a half point in 2 views. The first scenario, where we choose the anchor point as the one point, is illustrated here further. Anchor point is the origin (0, 0, 0) in the object coordinate system. Say, image coordinates of the anchor point in the first and second views are respectively (x_a, y_a) and (x'_a, y'_a). Substituting (X, Y, Z) with (0, 0, 0), (x, y) by (x_a, y_a), and (x', y') by (x'_a, y'_a) in (45), we obtain the following equations,

$$\begin{aligned} X_c &= -x_a , \\ Y_c &= -y_a , \\ \sin\beta . Z_s - (\cos\beta - 1) . X_s + T_x - x'_a + x_a &= 0 , \end{aligned} \tag{46}$$

where the last equation is simplified applying the first one, $X_c = -x_a$. From (46) we see that X_c and Y_c can be known from (x_a, y_a). To solve for Z_s and X_s, we need one more equation which we can obtained from a third view taken at another tilt angle. For tilt angles β_1 and β_2 we obtain the following two equations,

$$\begin{aligned} \sin\beta_1 . Z_s - (\cos\beta_1 - 1) . X_s + T_{x_{\beta_1}} - x'_a + x_a &= 0 \\ \sin\beta_2 . Z_s - (\cos\beta_2 - 1) . X_s + T_{x_{\beta_2}} - x''_a + x_a &= 0 \end{aligned} \tag{47}$$

Z_s and X_s are solved from the pair of equations in (47). x_a, x'_a, and x''_a are the x image coordinates of the anchor point in the first, second, and third views. Finally, we obtain the following reconstruction equations for (X, Y, Z) of a matched pair of points $\{(x, y), (x', y')\}$ in the first and second views (i.e. left and right stereo images) with relative tilt angle β,

$$X = X_c + x$$

$$Y = Y_c + y \text{, and}$$

$$Z = \frac{-\sin\beta.Z_s + (\cos\beta - 1).X_s - T_x + x' - x - (X_c + x).(\cos\beta - 1)}{-\sin\beta}. \quad (48)$$

For a constrained setup (Figure A2), the set of equations in (48) can be simplified in the following way. Assume that stage coordinate system is aligned with object coordinate system (X_s, Y_s, Z_s) = (0, 0, 0) and camera coordinate system differs only in Z. In this case, T_x=0, X_c=0, and Y_c = 0. Applying these values in (48), (X, Y, Z) is approximated into

$$X = x$$

$$Y = y \text{, and}$$

$$Z = -x.\tan\frac{\beta}{2} + \frac{x - x'}{\sin\beta}. \quad (49)$$

The expressions in (49) have been used in a paper by Hemmleb et at. [Hemmleb97] .

While imaging the stereo images, working distance of the SEM/LC-SEM system has to remain the same. When tilted, the object goes out of focus; the stage is then translated along its optical axis (which is parallel to the Z axes in our set up) to bring the object in focus. This movement changes the Z-location of the origin of the scanner coordinate system, (X_c, Y_c, Z_c). However, this change does not affect the reconstruction; since, in equations (48) or (49) we do not have Z_c.

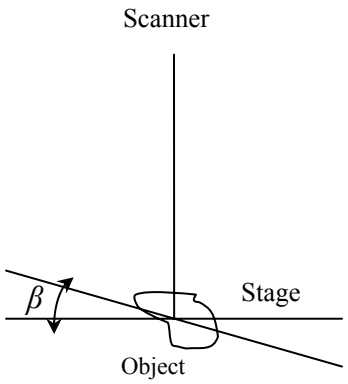


Figure A2: Constrained SEM/LC-SEM stereo imaging set up

Vita

Mohammad Shafikul Huq was born in Bangladesh. He completed his undergraduate study in the Department of Computer Science and Engineering from Bangladesh University of Engineering and Technology (BUET), Dhaka, Bangladesh, in 1997. Then, he joined the Department of Computer Science and Engineering in Ahasanullah University of Science and Technology (AUST), Dhaka, Bangladesh, as a lecturer. After two years of teaching in AUST, Mr. Huq came into the USA for higher study in 1999. He started his MS with major in Computer Engineering in the department of Computer Science and Engineering of Wright State University, Dayton, OH, with thesis work in computer vision under supervision of Dr. Ardeshir Goshtasby. Upon completion of MS in 2001, he was appointed as an R&D employee in Digital Optics Technologies, Inc., Rolling Meadows, IL, for a short period of time. Then, he came to The University of Tennessee at Knoxville in Summer, 2003, as a PhD student under Dr. Mongi A. Abidi. 2D/3D Image Processing and Computer Vision are his major backgrounds.