2010

Stereo Correspondence Based On Multiresolution Analysis

Dibyendu Mukherjee
University of Windsor

STEREO CORRESPONDENCE BASED ON MULTIRESOLUTION ANALYSIS

by

Dibyendu Mukherjee

A Thesis
Submitted to the Faculty of Graduate Studies
through Electrical and Computer Engineering
in Partial Fulfillment of the Requirements for
the Degree of Master of Applied Science at the
University of Windsor

Windsor, Ontario, Canada

2010

Stereo Correspondence Based On Multiresolution Analysis

by

Dibyendu Mukherjee

APPROVED BY:

Dr. Niharendu Biswas
Department of Civil and Environmental Engineering

Dr. Esam Abdel Raheem
Department of Electrical and Computer Engineering

Dr. Q.M. Jonathan Wu
Department of Electrical and Computer Engineering

TBA
Chair of Defense

7 May, 2010

Declaration of Previous Publications

This thesis includes 3 original papers that have been previously published/in press/ accepted for publication in peer reviewed journals and conferences, as follows:

Thesis Chapter	Publication title/full citation	Publication status
Chapter 4	D. Mukherjee, G. Wang, Q.M.J. Wu. "Stereo Matching Algorithm based on Curvelet Decomposition and Modified Support Weights"; IEEE International Conference on Acoustics, Speech and Signal Processing, ©2010 IEEE. Reprinted, with permission	published
Chapter 4,5	D. Mukherjee, G. Wang, Q.M.J. Wu. "Stereo Correspondence based on Curvelet Decomposition, Support Weights and Disparity Calibration"; SPIE Journal of Optical Engineering	in press
Chapter 6	D. Mukherjee, G. Bhatnagar, Q.M.J. Wu. "Performance Evaluation of Multiresolution Methods in Disparity Estimation"; International Conference on Image and Signal Processing	accepted

I certify that I have obtained a written permission from the copyright owner(s) to include the above published material(s) in my thesis. I certify that the above material describes work completed during my registration as graduate student at the University of Windsor.

I declare that, to the best of my knowledge, my thesis does not infringe upon anyones copyright nor violate any proprietary rights and that any ideas, techniques, quotations, or any other material from the work of other people included in my thesis, published or otherwise, are fully acknowledged in accordance with the standard referencing practices. Furthermore, to the extent that I have included copyrighted material that surpasses the bounds of fair dealing within the meaning of the Canada Copyright Act, I certify that I have obtained a written permission from the copyright owner(s) to include such material(s) in my thesis.

I declare that this is a true copy of my thesis, including any final revisions, as approved by my thesis committee and the Graduate Studies office, and that this thesis has not been submitted for a higher degree to any other University or Institution.

Co-Authorship Declaration

I hereby declare that this thesis incorporates material that is result of joint research, as follows:

This thesis also incorporates the outcome of a joint research undertaken in collaboration with Dr. Guanghui Wang and Dr. Gaurav Bhatnagar under the supervision of professor Jonathan Wu. The collaboration is covered in Chapter 4, 5 and 6 of the thesis. In all cases, the key ideas, primary contributions, experimental designs, data analysis and interpretation, were performed by the author, and the contributions of co-authors were primarily through the provision of proof reading and reviewing the research papers regarding the technical content.

I am aware of the University of Windsor Senate Policy on Authorship and I certify that I have properly acknowledged the contribution of other researchers to my thesis, and have obtained written permission from each of the co-authors to include the above materials in my thesis.

I certify that, with the above qualification, this thesis, and the research to which it refers, is the product of my own work.

Abstract

Disparity estimation from stereo imagery has gained substantial interest of research community from its commencement with the recent trend being the use of multiresolution methods. Existing multiresolution based methods generally rely on approximate band based matching neglecting other subbands that carry high-frequency information. Present research is an effort to find a multiresolution based stereo correspondence method that effectively uses the high and low-frequency subbands in multiple resolutions, bridges the gap between feature-based and area-based matching by incorporating the vast feature space of multiresolution and develops an optimum approach between global and local area-based methods in terms of accuracy, implementation and computational complexity. As a response to the lack of exploration of different multiresolution based stereo, a novel comprehensive comparison framework is proposed to evaluate different multiresolution based disparity estimation methods. Extensive qualitative and quantitative results with detailed analysis have been provided to support the claims of the work.

to my

mother and father

and my loving wife

Acknowledgements

I express my sincere gratitude to my advisor, Dr. Q.M. Jonathan Wu for giving me the opportunity to work under his supervision as well as for his guidance and support in my research. I would like to thank Dr. Esam Abdel-Raheem for his valuable suggestions and comments. I would also like to convey my sincere thanks to Dr. Nihar Biswas for his priceless advices and help. I would like to express my thanks to Dr. Boubakeur Boufama and Dr. Dan Wu for their support and help. I would like to take this opportunity to thank Dr. Guanghui Wang and Dr. Gaurav Bhatnagar for all the guidance and help in my research work.

I am thankful to my friends and colleagues, in particular, Tanaya, Rashid, Adeel, Pankaj, Aryaz, Thanh, Fei, Shilpi, Sanjib, Sarmistha, Himadri, Rajashri, Saptarshi and Iqbal for their help and frequent assistances throughout my work and stay. Last but not least, I would like to thank my family for their love and confidence in my abilities that helped me to move on.

Table of Contents

List of Tables

List of Figures

List of Acronyms

AdaptDispCalib	Local Stereo Matching with Adaptive Support-Weight, Rank Transform and Disparity Calibration
adaptWeight	Adaptive Support Weights
CONT	Contourlet Transform
CoopRegion	Region based Stereo Matching Algorithm using Cooperative Optimization
CostAggr+occ	Cost Aggregation and Occlusion Handling with WLS in Stereo Matching
CT	Curvelet Transform
Curv+MASW	Curvelet Decomposition with Modified Adaptive Support Weights
curv+MASW+DC	Curvelet Decomposition with Modified Adaptive Support Weights and Disparity Calibration
DC	Disparity Calibration
DoubleBP	Stereo Matching with Color-Weighted Correlation, Hierarchical Belief Propagation and Occlusion Handling
fixWin	Fixed Window based Sum of Absolute Differences

GC+occ	Computing Visual Correspondence with Occlusions using Graph Cuts
GeoSup	Local Stereo Matching using Geodesic Support Weights
MASW	Modified Adaptive Support Weights
MRSVD	Multiresolution Form of Singular Value Decomposition
OrigSAD	Original SAD based Stereo Matching
RealtimeBP	Real-Time Global Stereo Matching using Hierarchical Belief Propagation
RegionTreeDP	Region-Tree based Stereo using Dynamic Programming Optimization
SAD	Sum of Absolute Differences
SegTreeDP	A Fast Line Segment Based Dense Stereo Algorithm using Tree Dynamic Programming
segWin	Segmented Window based Sum of Absolute Differences
segWinApprox	Segmented Window based Sum of Absolute Differences on Approximate Band
SSD	Sum of Square Differences
SVD	Singular Value Decomposition
symBP+occ	Symmetric Stereo Matching for Occlusion Handling
Wave+MASW	Wavelet with Modified Adaptive Support Weights
WT	Wavelet Transform
ZNCC	Zero Mean Normalized Cross Correlation

Chapter 1

Introduction

Computer vision is a vast area of research. Among many interesting areas, the reconstruction of 3-D structure and shape from a number of 2-D images of a scene is still a challenging field for research. 3-D reconstruction is a process to recover the depth and structure of an object or a scene from images. Whenever a scene is captured in an image, the depth information of the scene is lost forever. Practically, there is no way to extract the depth information from a single image of the scene if the objects in the scene have no known surface regularity that can be modeled.

There exists a number of methods for depth recovery from multiple images of a scene. Stereo vision, shape from focus, shape from shading and shape from defocus are some examples. Stereo Vision is one of the most popular methods for depth estimation. The next section provides a brief overview of stereo Vision.

1.1 Stereo Vision

The simplest and popular method for depth recovery is stereo vision or stereo reconstruction that takes a pair of images of a scene and uses a technique commonly called *Triangulation* to extract the depth from the images. When two image points in the image pairs corresponding to a specific scene point are known, the depth of the scene point can be recovered. A complete scene reconstruction requires exact

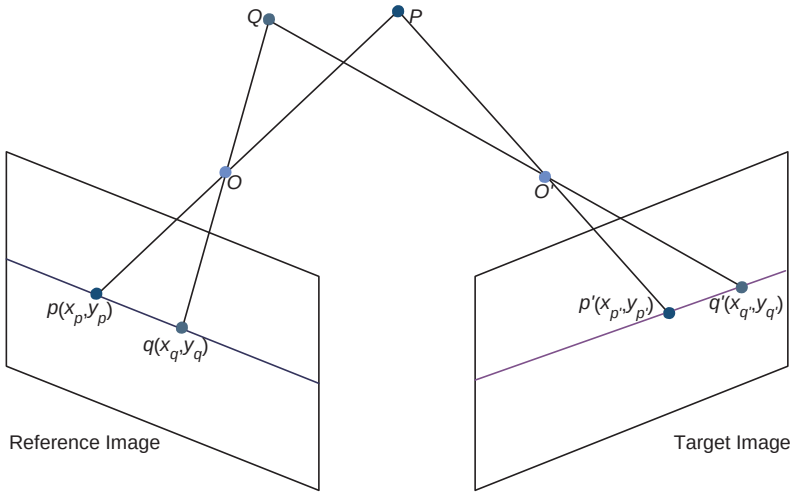

Figure 1.1: A scene point P is projected in reference and target images as p and p' respectively. Scene point Q is projected as q and q'.

matching of all image points between the two images. Generally, the image points of first or *reference* image are matched to the image points of the second image called *target* image and the matching process is termed *correspondence matching*. Correspondence matching yields the shift of position of the projection of a scene point from the reference image to the target image.

Figure 1.1 shows the process of triangulation. Scene points P and Q project to p and q in reference image, respectively. Similarly, they project to p' and q' in target image. A point $p(x_p, y_p)$ denotes that x_p and y_p represent the $X--$ and $Y--$coordinates of p. The camera centers are denoted by O and O' respectively. If the image point pair (p, p') is known, the projection lines $\overline{pO}$ and $\overline{p'O'}$ intersect in P which is uniquely estimated. The shift of location of an image point $p(x_p, y_p)$ in reference image to image point $p'(x_{p'}, y_{p'})$ in target image is termed as *Disparity* or *Correspondence*. Mathematically, it can be expressed as $d = (x_p - x_{p'}, y_p - y_{p'})$. The criteria for a depth recovery is that the scene point must be visible from both the cameras. If the correspondence pair does not exist, there is no way to extract the

depth information.

Note: Throughout the research document, the terms *disparity estimation* and *stereo correspondence* will be used interchangeably.

If a gray level image is constructed, whose pixel values are scaled disparity values of the image pair, the image is termed *Disparity Map*. Based on the number of correspondence pair matched, disparity map can be categorized as *dense* and *sparse*. A dense disparity map contains disparity for each pixel of the reference image, while, a sparse map contains disparities for a fraction of the total number of pixels. For a complete scene reconstruction, a dense disparity map is required. The following sections explain the challenges involved in a dense disparity estimation followed by the the motivation and the problem statement of the research.

1.2 Challenges

In simple terms, stereo Correspondence is matching the reference image points to the target image points. Apparently, matching two pixels seems easy. But there exist a number of factors that make stereo correspondence a highly challenging task. Some of the factors are discussed below.

1. **Occlusions and Discontinuities:** A scene is captured by two cameras from two different angles. This results in a shift of objects with respect to each other between the reference and target images. Near the object boundaries where two objects meet, some pixels get occluded by the object on the front. The amount of occlusion varies between views. Thus, an image point may be visible from one image while invisible in the other. This effect is called *half-occlusion* and generally occurs near object boundaries (discontinuity). Occlusion near discontinuity is shown in Figure 1.2.

   In Figure 1.2(a), the chimney of the toy house is shifted to left in the target image. The image points in the background visible from the reference image

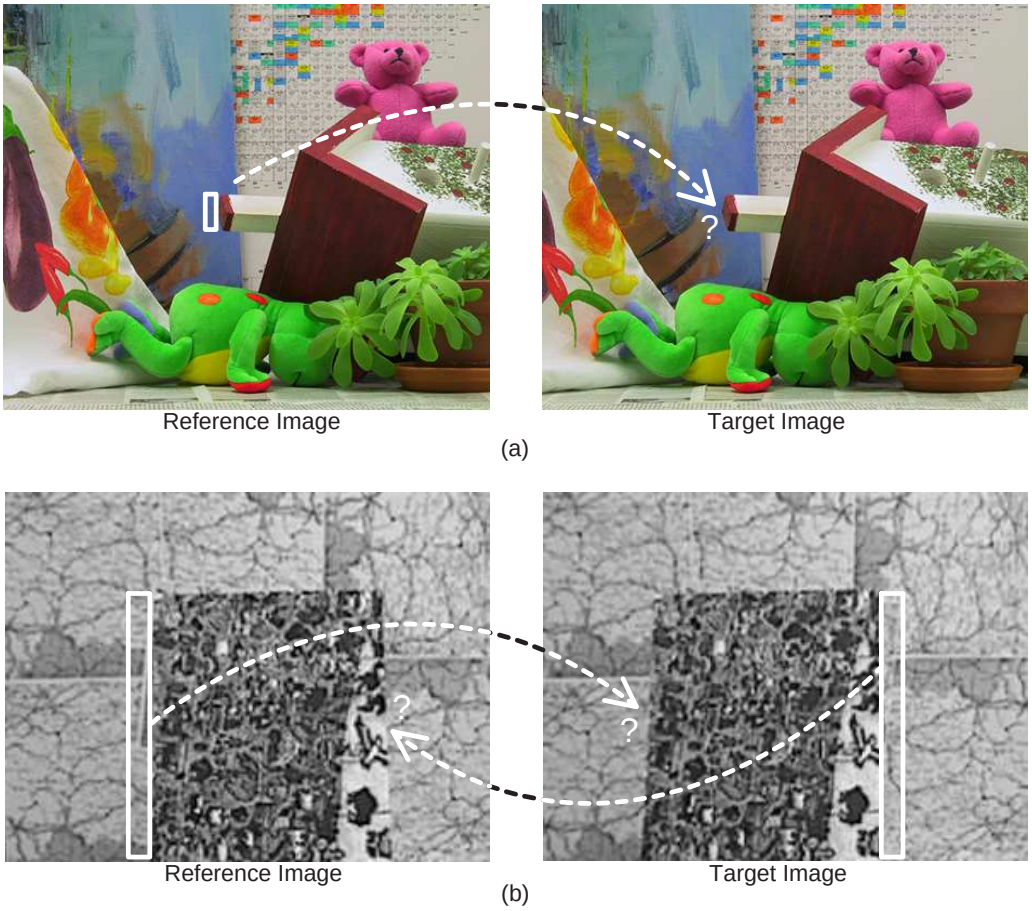

Figure 1.2: Occlusions in stereo images. (a) The background is visible inside the white rectangle in reference image but occluded by chimney in target image. (b) The background is half-occluded in both images.

are occluded by the chimney in the target image. These points do not have any correspondence pair and cannot be matched by searching. Similarly, in Figure 1.2(b), the background is occluded on the right in reference image and on the left in target image. But, in both cases, the correct matching is not possible.

2. **Repetitive Patterns and Uniform Patterns:** Repetitive patterns repeat themselves in a region, while, the regions with uniform patterns consist of many

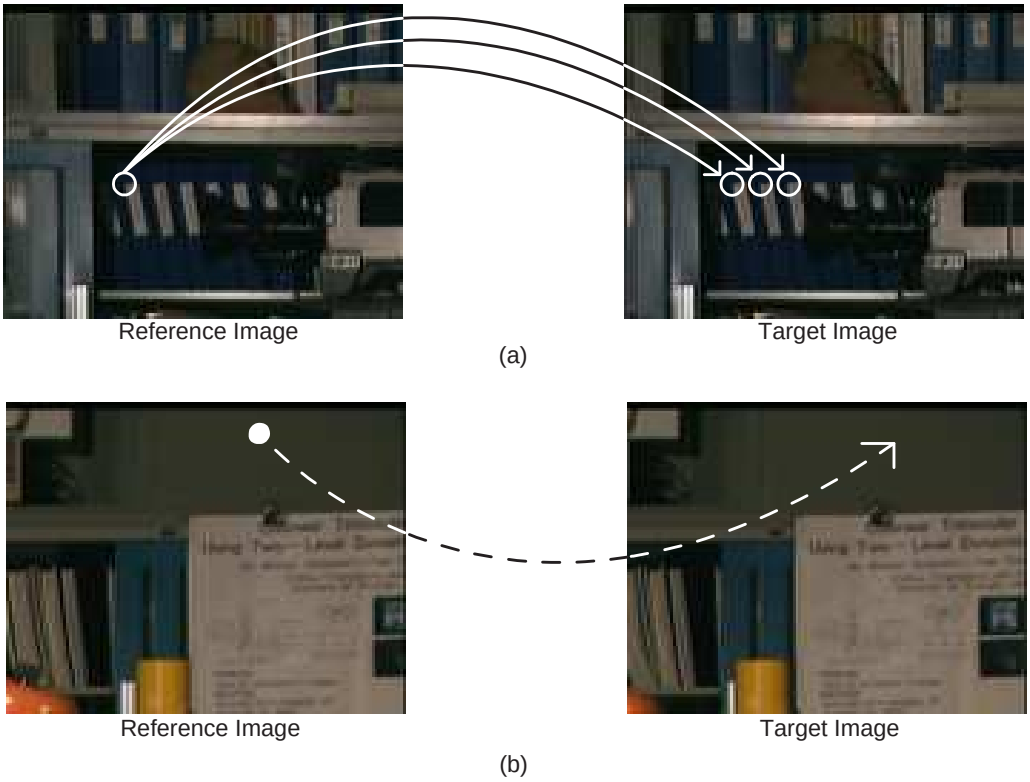

Figure 1.3: Repetitive and uniform patterned regions. (a) The folders at the back have repetitive pattern resulting in ambiguous matching. (b) The background image points have uniform gray values resulting in multiple choices for matching.

pixels with similar color and intensity profile. They provide ambiguous choices for matching a candidate.

Any repetition has atleast another candidate with similar properties to be matched. The problem is shown in Figure 1.3(a) which shows the top-left regions of the Tsukuba pair. As shown in the figure, the folders at the back have similar patterns. The circles represent the prospective candidates for matching. As the candidates are almost identical, the matching is ambiguous.

Likewise, regions with uniformity in patterns are difficult to match. The top-

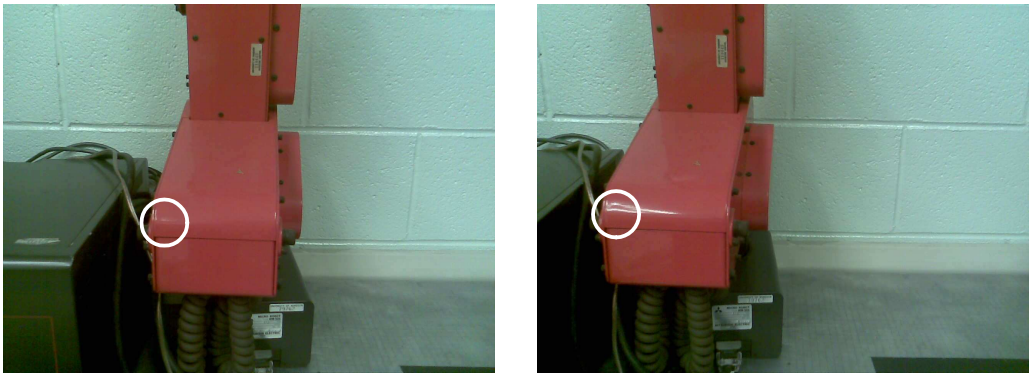

Figure 1.4: The machine has metallic surface which is specular in nature. The target image has a reflection in the circled region.

right regions of Tsukuba pair as shown in Figure 1.3(b) demonstrate the problem. The background pixels in the images have similar color and pattern. Thus, for each pixel in reference image, there exist a number of potential candidate pixels in target image.

3. **Specularity:** A specular surface reflects light from a single incoming direction into a single outgoing direction abiding by the laws of reflection in contrast to the diffused reflection, where, the light is reflected in a broad range of directions. In photometry, specular surface is also known as *non-Lambertian* surface, because, a *Lambertian* surface scatters the incoming light such that the brightness of the surface remains same from any direction of observation. Some examples of specular surface are glass, stainless steel or any shiny material. Figure 1.4 shows the difficulty with specular surfaces.

   The metallic object has specular surface which can reflect light in a specific direction collinear with the target-view direction. Due to the reflection, the circled region cannot be matched properly.

4. **Perspective Distortion and Foreshortening:** The projective geometry of

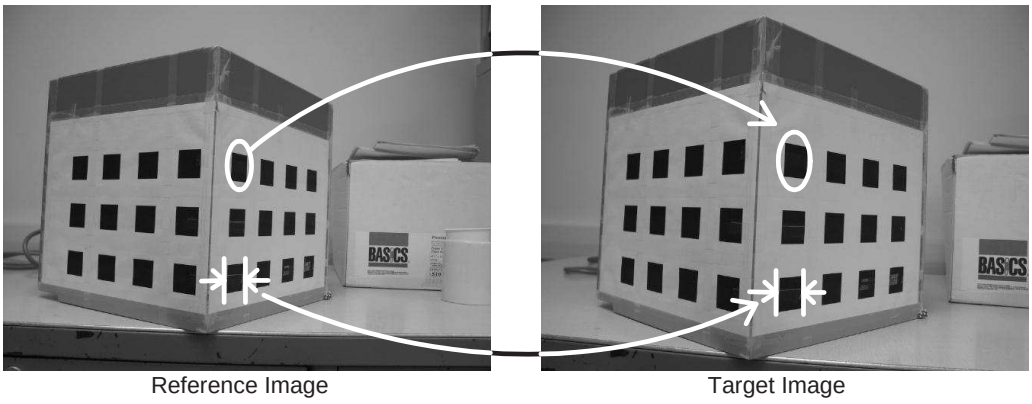

Figure 1.5: The two images are taken from different perspective views. The circled area in reference image is smaller and distorted compared to that of target image due to larger distance from camera. The width of the black square is also foreshortened in reference image.

3-D scenes make them perspective in nature. The properties of perspective view are as follows:

- The object becomes smaller as the distance from observer increases
- The length of object along the line of sight is relatively shorter than the length across the line of sight. This phenomenon is called *Foreshortening*.

The effects of perspective distortion and foreshortening are depicted in Figure 1.5. The circled area is distorted due to the apparent change in perspective view. Thus, the total number of pixels to match are different in two views of the circled region. The foreshortened width of the black square makes it difficult for correspondence matching.

5. **Photometric Distortions and Noise:** Photometric distortions and noises are inherently present in any capture by a camera. The two images have lighting variations due to the change of camera positions, camera noise etc. Change in intensity level is very common between two views. Figure 1.6 shows the

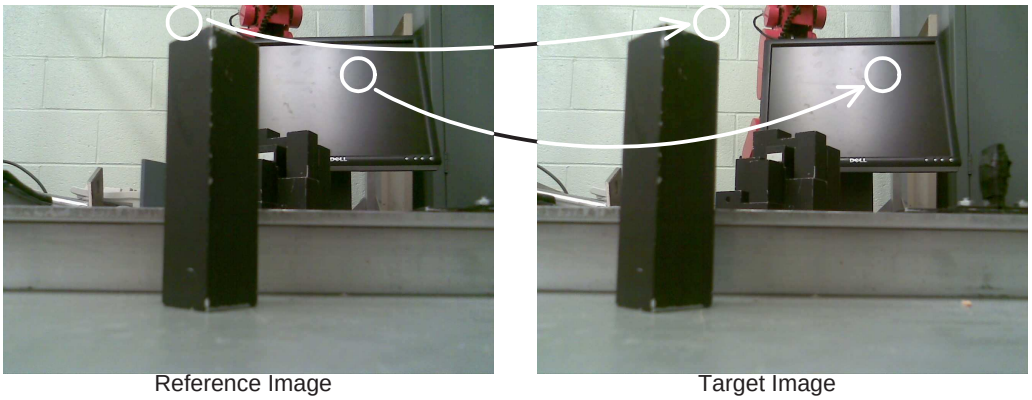

Figure 1.6: The circled regions have different illumination variations due to the change in camera view.

problem.

The circled regions in the images have photometric distortions. A key matching criteria is intensity profile. Due to the changes in intensity in the circled regions, the matching is very difficult.

6. **Transparent Objects:** Transparency presents another difficulty in matching. A transparent object surface contains the background as well. A shift of the view may create a background mismatch, and consequently, correspondence mismatch. The problem is shown in Figure 1.7. The circled areas are on the bottle. As the bottle is transparent, the background is partially visible. With the shift in location, the background is changed and the change can be seen inside the circled region in the target image. The transparency of the material makes correspondence matching more difficult. It is also worthy to note that the transparent objects suffer from specularity problem as can be seen from the figure.

The aforesaid challenges together make stereo correspondence estimation a difficult problem in computer vision. The extensive searching operations involved in the pro-

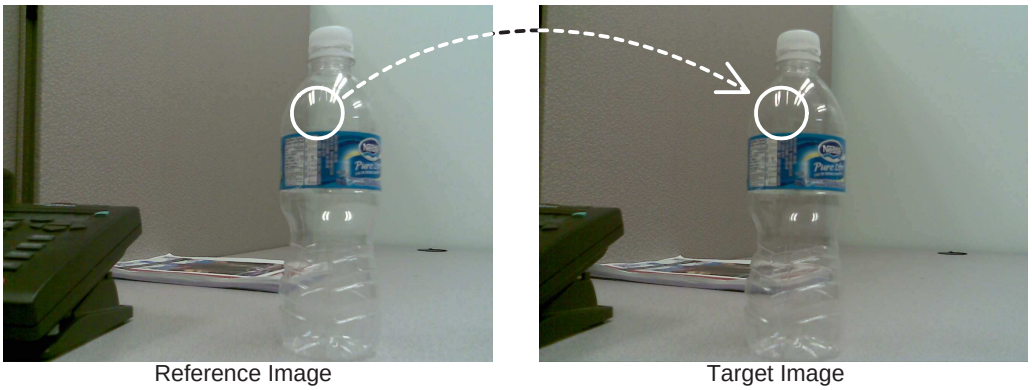

Figure 1.7: The background in the circled areas are different but they are not totally occluded by the transparent bottle.

cess also makes it computationally expensive. Thus, even with the current progress in computer vision, stereo correspondence is an ill-posed problem.

1.3 Motivation

The challenges in stereo correspondence estimation as described in Section 1.2, make it one of the fundamental subject of investigation in computer vision. There has been numerous research on this subject. An elementary but important question can be raised - does it worth the amount of research involved? The answer to the question would lie in the number of application areas. Stereo Vision is used primarily for depth estimation and has been in use along with range imaging or range sensors. Range sensors are famous for easy depth estimation but not suitable for all purposes. They are difficult to use with tiny robots. Active sensors affect other devices. They also get affected by other IR and LASER rays. On the other hand, stereoscopy is easier to apply in these sectors. Stereoscopic sensors are immune from spurious readings, cross-talks and multiple reflections present in IR, LASER or SONAR sensors. They are also low cost, realtime, robust and suitable for any size of vehicles. Naturally,

considerable amount of research has been done to improve the robustness of stereo algorithms.

Currently, researchers try to find better solutions to the challenges. Some of the state of the arts methods are listed with their results in the Middlebury stereo evaluation webpage [1]. A tradeoff exists between accuracy and speed. The methods with higher accuracy suffer from higher implementation and time complexity, whereas, fast and low complexity methods are often of lower accuracy. Yet, there exists no method that can completely satisfy atleast one of the challenges. The motivation of the work comes from this aspect of stereo correspondence analysis.

1.4 Problem Statement

The goal of this research is to find a method that can work with higher accuracy in low computation and implementation complexity. Searching for a possible solution to all the problems present in disparity estimation would be an intemperate challenge and probably infeasible. Instead, the focus of the work is to provide a generic solution that can bridge gaps among higher accuracy, higher speed and lower implementation complexity, and also partially defend against occlusions, repetitive and uniform patterns that are the most common problems in stereo. The work is based on multiresolution analysis that provides pyramidal structure to optimize searching and a high number of image features for better matching accuracy.

1.5 Objective

The research is targeted towards improvement in accuracy of stereo correspondence using multiresolution analysis and comparison among different multiresolution based correspondence methods. It involves the development of a multiresolution based disparity estimation method that yields good accuracy in lower implementation com-

plexity. The study also aims to improve the accuracy of disparity estimation with the use of continuity constraints. The objective also includes a comparison and performance evaluation of different multiresolution based stereo correspondence estimation methods.

1.6 Scope of this Work

A novel approach towards disparity estimation based on multiresolution analysis is discussed in the thesis. This work includes the following studies:

1. Stereo correspondence estimation using curvelet decomposition and modified adaptive support weights.

2. Improvement on correspondence estimation using curvelet decomposition, adaptive support weight and disparity calibration.

3. A proposed novel framework for evaluation of different multiresolution methods for stereo correspondence with comparisons and evaluations for wavelets, curvelets, multiresolution form of singular value decomposition and contourlets for disparity estimation.

1.7 Organization of Thesis

The rest of the thesis is organized as follows. In Chapter 2, a detailed review of the related works is provided with the general workflow, constraints and types of stereo correspondence methods. Chapter 3 gives an overview of the different multiresolution methods used in the research. Chapter 4 and 5 explain the proposed methods with experimental results and comparison tables. Chapter 6 introduces a novel framework to evaluate the performances of different multiresolution based correspondence methods. Chapter 7 summarizes the contributions and provides scope for future work.

Chapter 2

Literature Review

Stereo correspondence algorithms search the disparity of the reference image pixels with respect to the target image pixels. There have been numerous researches for accurate disparity estimation. Current state of the arts methods have reached high accuracy but with high computational and implementation complexity. In this chapter, the general workflow of any correspondence estimation method is described. The workflow is followed by a brief description of the basic constraints of stereo vision. The review also describes the broad classification of algorithms followed by the individual descriptions.

2.1 General Workflow

In literature, most of the correspondence algorithms follow some basic steps. According to the taxonomy in [31], a general workflow is discussed below.

1. **Matching Cost Computation:** In this step, a dissimilarity measure is employed to compute pixel-wise matching costs. Examples of such measures are *Euclidean distance* (squared difference), *Color difference* (absolute or square difference in color space) and *Manhattan difference* (absolute difference). The computed matching costs for each pixel for all disparities are the outputs of the

step.

2. **Cost Aggregation:** In this step, the computed matching costs are summed or averaged over a support region in the image space. The support region chosen may be two-dimensional for a fixed disparity to favor fronto-parallel surfaces or three-dimensional in the 3-D space constituted by image dimension (x, y) and disparity values d (xyd space) to favor slanted surfaces. Two dimensional regions inherently assume the pixels to be from same depth resulting in an unwanted fattening effect. This effect is explored further in Chapter 4. Examples of 2-D support regions are square windows, shifted windows, segmented windows and Gaussian convolution. Some examples of 3-D regions are limited disparity difference [14] and limited disparity gradient [29].

3. **Disparity Computation/Optimization:** This step is one of the most important step of the workflow that generates the disparity map. In this step, disparity is chosen or computed using the aggregated costs. The type of computation divides the algorithms into two categories - local and global. Local algorithms mainly compute pixel-based disparity and they can simply choose the disparity associated with the minimum cost for each pixel. Global algorithms, on the other hand, optimizes the disparity plane by some energy minimization procedure. Local and global algorithms are discussed in section 2.3.1 and 2.3.2 respectively. The output of the step is the initial disparity map. If the last step is not performed, this may be the final disparity map.

4. **Disparity Refinement:** The final step refines the disparity map by some post-processing methods like iterative optimization, curve fitting, mean or median filtering, left-right consistency check or by application of constraints (discussed in section 2.2). The output of this step is the final disparity map.

2.2 Constraints

Stereo vision is essentially a two view search problem. Each pixel in the reference view can have one matching pixel in the target view at the most. But, the pixel can lie on any row or column of the two dimensional image. An exhaustive two dimensional searching operation needs to be performed to find out the exact location of the matching pixel. Due to the challenges mentioned in section 1.4, stereo-matching is still a very complex and time consuming process. Fortunately, due to the projective geometry in stereo vision and properties of natural images, there exists a number of constraints that reduces the complexity of matching and provides higher accuracy in search. The constraints along with the relative geometry are described next.

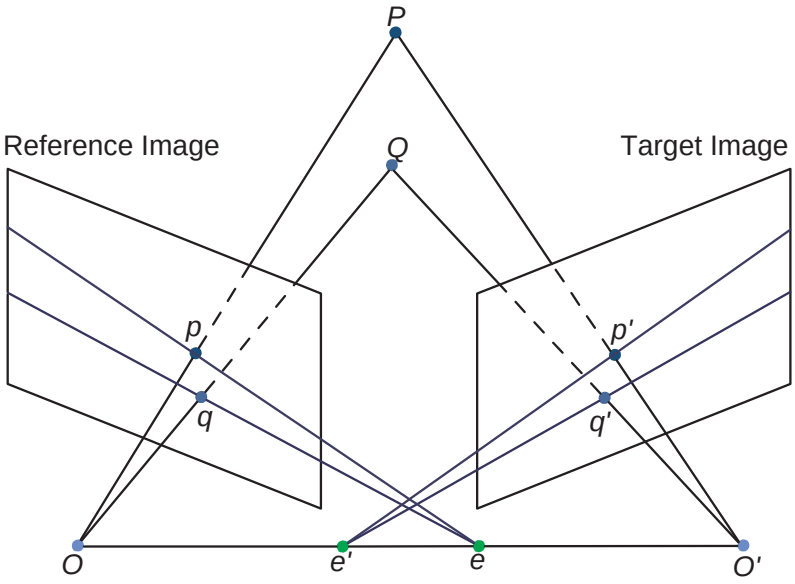

Figure 2.1: A scene point P is projected in reference and target images as p and p' respectively. Scene point Q is projected as q and q'.

1. **Epipolar Constraint:** Consider, two scene points P and Q are projected onto two image planes as shown in Figure 2.1. P and Q have projections p, p' and

q, q' respectively. For clarity, the camera centers O and O' are moved to the back of the camera plane. This does not make any difference to the computation, but simplifies the diagram to a large extent. The points O, O', p and p' are coplanar because O,O' and P are coplanar. Thus, the cross product of $\overrightarrow{O'O}$ and $\overrightarrow{Op}$ is a vector perpendicular to the plane defined by O, O', p and p' and hence, perpendicular to $\overrightarrow{O'p'}$. Mathematically, the following can be written:

$$\overrightarrow{O'p'} \cdot (\overrightarrow{O'O} \times \overrightarrow{Op}) = 0, \tag{2.1}$$

where $\cdot$ and $\times$ denote the dot and cross products, respectively. However, this relationship holds only if all the vectors are defined in same reference frame. Taking the target camera plane as reference, Eq. 2.1 becomes

$$\overrightarrow{O'p'} \cdot (\overrightarrow{O'O} \times R\overrightarrow{Op}) = 0, \tag{2.2}$$

where R is the 3×3 rotation matrix from the reference frame of target camera to that of reference camera. There also exists a translational part, but due to the cross product, it becomes zero. Let's denote $\overrightarrow{O'p'}$,$\overrightarrow{O'O}$ and $\overrightarrow{Op}$ by p'^T,t and p respectively. Then, Eq. 2.2 becomes

$$p'^T \cdot (t \times Rp) = 0, \tag{2.3}$$

where $t = (t_x, t_y, t_z)$ represents the coordinates of O in target camera reference frame. Again, $t\times$ can be written in a rank-deficient matrix form S as

$$S = \begin{bmatrix} 0 & -t_z & t_y \\ t_z & 0 & -t_x \\ -t_y & t_x & 0 \end{bmatrix}. \tag{2.4}$$

Thus, Eq. 2.3 becomes

$$p'^T SRp = 0. \tag{2.5}$$

Normally, the product SR is represented as a matrix E called the *essential matrix*. Using the notation, Eq. 2.5 can be written as:

$$p'^T Ep = 0. \tag{2.6}$$

Geometrically, Ep is a product of a 3×3 matrix and a 3×1 vector. Thus, Ep can be represented as a 3×1 vector as shown

$$V = \begin{pmatrix} a \\ b \\ c \end{pmatrix}.$$

Also, P'^T can be expressed in projective space as $(x, y, 1)$ because it has zero $Z - -$component. Then, $p'^T V$ becomes

$$ax + by + c = 0, \tag{2.7}$$

which states that point p' on target image lies on the line defined by Ep. This line is called the *epipolar line*. By geometry, the corresponding point p' is constrained to lie on the epipolar line defined by the reference point p and the essential matrix E. This leads to the following points:

- The epipolar line Ep is a representation of the vector $\overrightarrow{Pp}$ in the target image.
- As all the vectors $\overrightarrow{Pp}$ in the reference image pass through camera center O, all the epipolar lines in the target image pass through the corresponding point of O in target image.
- The corresponding point of O in target image is called the *epipole* e. As stated above, all epipolar lines pass through e. This can be observed from Figure 2.1.

Thus, epipolar line reduces the burden of searching for a corresponding point in two-dimensional space to a one-dimensional line.

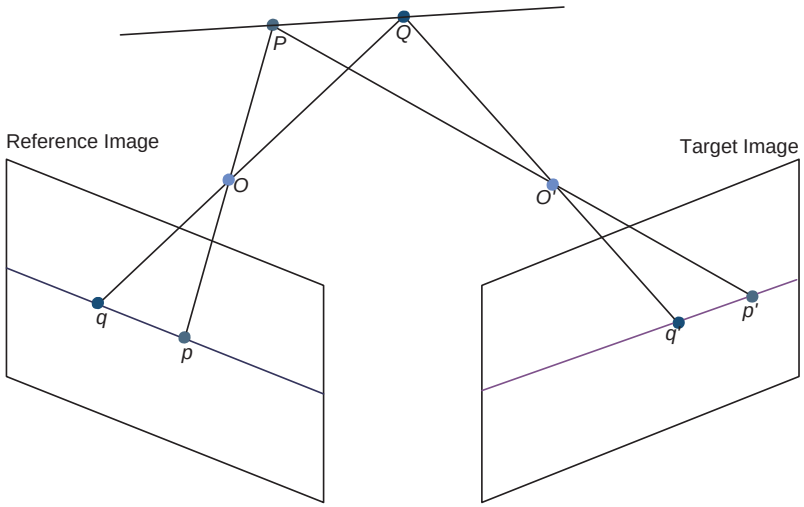

Figure 2.2: The relative order of p and q on the reference image epipolar line, are maintained in the target image epipolar line for p' and q'.

2. **Order Constraint:** The order constraint states that the image points lying on the reference epipolar line appear in the same order in the target epipolar line. This constraint is clear from Figure 2.2.

3. **Uniqueness Constraint:** The uniqueness constraint comes from the fact that a scene point can have a single projection in one image. Thus, each image point in reference image can have maximum one corresponding point in target image. In fact, there may not exist a corresponding point at all.

4. **Continuity Constraint:** Objects have continuous surfaces that provide depth continuity along the surface. Depth discontinuity occurs at object boundaries in general. The continuity constraint provides a measure to predict the depth (disparity) of a image point on an object surface from the depths (disparities) of the previous image points on the same object surface. The continuity constraint has been successfully used in the proposed method described in Chapter 5.

The constraints discussed above help to reduce false matches, reduce searching complexity and provide better prediction for correct matches.

2.3 Method Classification

There are presently two types of approaches for stereo correspondence matching–area-based and feature-based.

Feature-based methods work by extracting local features from the images and matching the features across the stereo pairs. The features, such as edges, corners, and lines, are matched using local feature descriptors. Local features remain more or less unaltered across the image pairs. So, feature-based methods are robust, accurate, and fast, although they usually generate sparse disparity maps that are not appropriate for some operations like dense reconstruction. There are broadly three types of features: interest points [26, 33], edges [34], and regions [27]. But there are also global feature based methods [40]. Global feature based methods, also termed *structural matching*, match larger features that consist of local smaller features.

Area-based approaches rely on some statistical correlation of color or intensity values. They can be classified into two types: local and global methods. The local methods estimate the disparity of a pixel by correlating a support window around the pixel with a similar support window in the other image. Area-based methods assume all pixels in the support window to have similar depth. This assumption is violated in depth discontinuities and results in a fattening effect near these regions. The methods select the best correlation match for each pixel. This sometimes results in wrong disparity estimation for image points having ambiguity of depth. As a consequence, textureless regions, repetitive textured regions, and regions with depth discontinuity fail to match correctly.

Some algorithms proposed for densely textured regions are variable windows [38] and shiftable windows [21]. Adaptive support window methods [4, 19] try to find

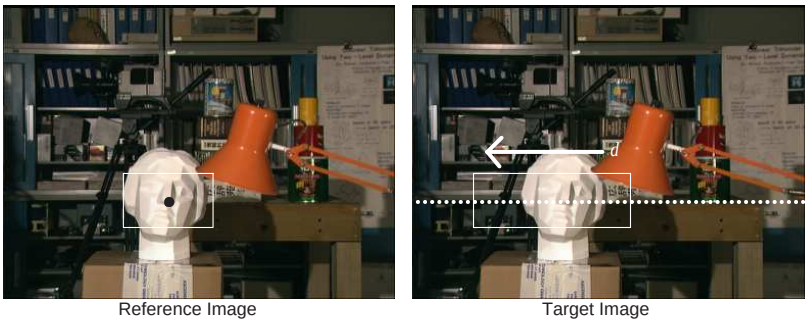

Figure 2.3: The support window in reference image is matched to a similar window in target image. The maximum shift of window is denoted by d. The epipolar line is shown as a dotted line in target image.

an optimum size support window for each pixel. An initial disparity based adaptive support window selection method was proposed by Kanade and Okutomi [19]. Veksler [37] proposed a method to compare different sized windows based on window cost. Multiple window methods try to select an optimal support window size from a predefined range of windows [3, 13, 20].

In contrast to local window-based methods, global area-based methods try to estimate an optimum disparity plane that minimizes a global cost function. Some global methods proved to be very effective, like cooperative optimization [41], belief propagation [42, 43], graph cuts [23], dynamic programming [24, 39], etc. Brief descriptions to local and global methods are presented next.

2.3.1 Local Methods

Local matching methods generally compute disparity map using step 1,2 and 3 in section 2.1. As shown in Figure 2.3, for matching a pixel shown in black dot, a support window around the pixel is considered. The support window is matched against a similar support window around pixels that lie on the epipolar line shown as a dotted line in target image.

A number of different matching cost functions are used by local methods, namely Sum of Absolute Differences (SAD), Sum of Square Differences (SSD) and Zero Mean Normalized Cross Correlation (ZNCC). Due to simple operation, local methods are very fast. But, the support window implicitly assumes that all pixels belong to same depth and the surface of the window is a fronto-parallel surface. Both the assumptions are incorrect in practical situations. In object boundaries, there exist depth discontinuities because two objects may lie in different depths. Also, object surfaces are not always fronto-parallel, but slant. Thus, the local methods have lower accuracy.

2.3.2 Global Methods

Global methods try to estimate the disparity plane using a global cost function. Generally, the cost function has the following form:

$$E(d) = E_{data}(d) + \lambda E_{smooth}(d), \tag{2.8}$$

where d is the disparity, E_{data} is known as the data cost and E_{smooth} is known as the smoothness cost. The data cost is generally the local matching cost like SAD, SSD etc. The smoothness cost is the cost of assigning disparities to neighbor pixels. If the smoothness cost is high for some disparity d, then it implies d violates the depth continuity around its neighboring pixels and should not be assigned.

The minimization of such a function $E(d)$ can generate a highly accurate disparity map. But, this minimization is a very complex process and may be computationally expensive. This disadvantage of global methods sometimes prevent them from being used in practical situations where low computational and execution complexity is desired.

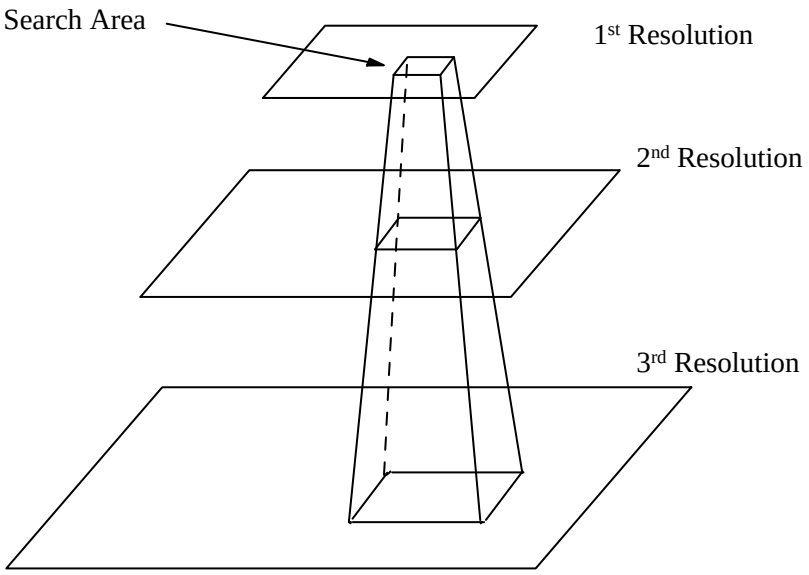

Figure 2.4: The smaller search area in lower resolution is extended in larger resolutions.

2.3.3 Multiresolution Methods

Multiresolution methods are relatively new members in the family of disparity estimation methods. They were introduced for stereo by Mallat [25] and Kim et al. [22]. Later, improved algorithms have been proposed by M. Shim [35], Zhang et al. [45] and Caspary and Zeevi [9]. Recently, Bhatti and Nahavandi [2] have proposed an effective way of stereo-matching based on multiwavelets that actually involves different detail subbands. Ding *et al.* [11] have proposed the use of shift invariant contourlet transform in stereo.

Classical approach of multiresolution based stereo is explained in Figure 2.4. As shown in figure, the image is decomposed into different levels of resolution. The disparity is estimated in the lowest or approximate resolution. This disparity is properly scaled and projected onto the next resolution and refined in neighboring region. The refinement is possible because higher resolution provides higher amount of details.

This iterative refinement of disparity is continued till the highest resolution i.e the original image. This progressive searching operation reduces number of comparisons. This can be explained as follows -

For simplicity, wavelet decomposition is taken as the multiresolution method. Let, the image has dimension $N \times N$ where, N is a positive integer.

If the image is decomposed into multiple resolutions, the dimension reduces by a factor of 2 with each level of decomposition. Thus, if the image is decomposed into n levels, the dimension of n^{th} level image is $(N/2^n) \times (N/2^n)$.

Let, the maximum disparity is d. Then, number of comparisons for n^{th} level image is $((N/2^n) \times (N/2^n)) \times d$ as each pixel in reference image is compared with d pixels in target image.

The initial disparity map is progressively corrected in higher resolutions by taking a small region around the initial projected disparity and searching for the best match inside the region. Let, each level consists of p extra searching for disparity refinement. Without violating the conditions, it can be assumed that $p \ll d$ as the refinement region is very small compared to the disparity d. Thus, total number of comparisons is

$$((N/2^n) \times (N/2^n)) \times (d + n \times p).$$

Simplifying the terms, the expression becomes

$$((N \times N) \times (d/2^{2n} + (n \times p)/2^{2n})).$$

As previously assumed, $p \ll d$. Again, n represents the level of decomposition. Usually, 2 to 4 level of decomposition is used. d can range from as low as 4 to as high as 200. Thus, Without any loss of generality, it can be assumed that $n < d$. This leads to the inequality

$$((N \times N \times n \times p)/2^{2n}) \leq ((N \times N \times d)/2^{2n}).$$

If both term are taken equal for simplicity, then the expression becomes

$$2 \times (N \times N \times d/2^{2n}).$$

Summing the powers of 2, the expression becomes

$$N \times N \times d/2^{2n-1}.$$

If multiresolution is not used, the dimension does not reduce and number of matching per pixel remains d. Thus, the total number of comparisons becomes

$$(N \times N) \times d.$$

As $2^{2n-1} \gg 1$, it is evident that

$$(N \times N) \times d \gg N \times N \times d/2^{2n-1}.$$

This proves the executional advantage of multiresolution based analysis over others.

The disadvantages of classical method lies in the fact that the approximate band decomposition does not contain the detail features of the image. The absence of important features degrades the disparity estimation process. This problem is well-handled if high frequency detail bands are included in the process. Also, based on the type of multiresolution method used, the quality of disparity map changes. These scenarios are explored in the work.

Chapter 3

Multiresolution for Stereo Vision

Searching disparity in multiple resolutions has been an interesting topic among researchers. Multiresolution methods bridge a gap between local and global methods in terms of execution speed and accuracy of disparity map generation. They are faster than global methods due to the searching space reduction as shown in section 2.3.3, and, more accurate than local methods due to higher number of matching features and progressive refine of disparity in multiple resolutions. Still, there has been less research reported on this field to enhance the performance of classical methods, or compare different multiresolution based disparity estimation methods. With a target to contribute to this research and gain more understanding of relative performances, the current work comprises of the development of two multiresolution based correspondence methods and a comparison among four different multiresolution methods for disparity estimation. This chapter briefly discusses the multiresolution methods used in the research.

3.1 Wavelet Transform

The necessity of wavelets came from the shortcomings of Fourier transform. Fourier transform is used to generate a frequency domain description of a signal. For a time-limited signal, it can provide the frequency components necessary to synthesize the

signal within the time interval. However, Fourier transform stretches the frequency response as the time interval reduces. This makes it impossible to understand which frequency components are the constituents for an arbitrarily small section of the total signal. This phenomenon is called the *Uncertainty Principal*, and, wavelets try to solve it as much as possible. Wavelets are time-limited functions that are localized in frequency [25]. For their transient nature, they are fundamentally different from the functions used for Fourier Transform. The 1D wavelet function ψ in continuous form is defined as

$$\psi_{a,b}(t) = \frac{1}{\sqrt{a}}\psi\left(\frac{t-b}{a}\right), \tag{3.1}$$

where a and b are the dilation (scale) and translation (position) parameters respectively. The 1-D continuous Wavelet Transform (WT) of a function $f(t)$ is defined as

$$W_f(a,b) = \int_{-\infty}^{\infty} f(t)\psi_{a,b}(t)dt. \tag{3.2}$$

Coming to the discrete domain, if the parameters a and b are represented as $a = a_0^j$ and $b = kb_0a_0^j$, then the wavelet in discrete domain may be obtained as

$$\psi_{j,k}(t) = a_0^{-j/2}\psi(a_0^{-j}t - kb_0), \tag{3.3}$$

where in general, $j, k \in Z$ and $a_0 = 2, b_0 = 1$. The 1-D discrete WT of a function $f(t)$ decomposes the function into two parts. The first part called the approximate part consists of the low frequency components of the function. The second part called the detail part consists of the high frequency components of the function. These parts are expressed as

$$W_\phi(j_0,k) = \frac{1}{\sqrt{M}}\sum_t f(t)\phi_{j_0,k}(t), \tag{3.4}$$

$$W_\psi(j,k) = \frac{1}{\sqrt{M}}\sum_t f(t)\psi_{j,k}(t), \tag{3.5}$$

where ϕ is called the scaling function with the same definition as of ψ. ϕ and ψ give the approximate and detail part respectively. Normally, j_0 is chosen as 0, M

as a power of 2 i.e, $M = 2^J$. Thus, $j = 1, 2, ..., J-1; k = 0, 2^0, 2^1, ..., 2^j$ retaining their previous definitions. As the WT is separable, the 2-D WT can be obtained using repeated 1D WT along $X-$ and then $Y-$axis. 2-D WT is used to analyze two dimensional signals.

2-D WT of an image decomposes it into an approximate subband and a detail part consisting of horizontal, vertical and diagonal subbands. The four subbands together produce a large number of features for matching across the stereo image pairs. For the current work, Daubechies wavelets have been used as a representative for the wavelet family.

3.2 Curvelet Transform

Wavelets as a multiresolution analysis method has had much success. Wavelets are capable of interpreting a signal as a sum of contributions in different scales and locations. They can detect point singularities in single dimensional signals due to their effective localization and multiscaling. But for higher dimensions, they yield poor results due to the lack of orientation selectivity.

Curvelets were developed by Candes and Donoho in 1999, mainly for image analysis. They have strong directional characteristics and due to their variable width and length with a parabolic scaling of length$^2\sim$width, the coefficients are highly anisotropic at fine scales.

As described by Candes [6], curvelets can be thought of as obtained by applying parabolic dilations, rotations, and translations to a specifically shaped function ψ; they are indexed using scale $a(0 < a < 1)$, location b, and orientation θ as

$$\psi_{a,b,\theta}(x) = a^{-\frac{3}{4}}\psi[(D_a R_\theta(x-b))] \tag{3.6}$$

with

$$D_a = \begin{pmatrix} \frac{1}{a} & 0 \\ 0 & \frac{1}{\sqrt{a}} \end{pmatrix},$$

where, D_a is a parabolic scaling matrix and R_θ is a rotation by θ radians. A close example of such a function would be multiscale ridgelets. The ridgelet theory was proposed by Candes. The involvement of ridgelets and formation of curvelets are well discussed in his paper [8].

Second-generation Curvelet Transforms (CTs) are simpler and faster than their previous versions. There are two types of implementation–via unequally spaced fast Fourier transform (USFFT) or via wrapping [7]. Curvelets via wrapping has been used here because it is faster and can be applied to any size of images effectively. For a 2-D function $f[t_1, t_2]$ with $0 < t_1, t_2 < n$ (a specific length), the main steps for wrapping are as follows:

1. Take the 2-D fast Fourier transform (FFT) of $f[t_1, t_2] \rightarrow \hat{f}[n_1, n_2]$.

2. Separate $\hat{f}[n_1, n_2]$ into dyadic subbands using a scale window W_j, with j representing the j$^{\text{th}}$ scale.

3. Separate each subband into angular wedges using angular windows $W_{j,l}$, with l representing l$^{\text{th}}$ wedge.

4. Form the product $\tilde{U}_{j,l}[n_1, n_2]\hat{f}[n_1, n_2]$, where, $\tilde{U}_{j,l}[n_1, n_2]$ is a discrete localizing window like in eq. (3.6).

5. Wrap the product inside a rectangle W of size $L_{1,j} \times L_{2,j}$ (in east-west) or $L_{2,j} \times L_{1,j}$ (in north-south) around the origin to obtain $\tilde{f}_{j,l}[n_1, n_2] = W(\tilde{U}_{j,l}\hat{f})[n_1, n_2]$, where, $L_{1,j} \sim 2^j$ and $L_{2,j} \sim 2^{\frac{j}{2}}$.

6. Take the 2-D inverse FFT of each $\tilde{f}_{j,l}$ to obtain curvelet coefficients at scale j and orientation l.

Generally, CT measure the information of a signal at specified scale and locations along specified orientations. Thus, they effectively represent objects with *curve-punctuated smoothness*–smoothness except discontinuity along a general curve with

bounded curvature [8, 6]. Real images have lots of curves, i.e., edges along different orientations. Using CT, these features can be represented more accurately compared to wavelets. Roughly, to represent an edge to a square error of $1/N$ requires $1/N$ wavelets but only $1/\sqrt{N}$ curvelets.

3.3 Multiresolution Form of Singular Value Decomposition

Multiresolution Form of Singular Value Decomposition (MRSVD) is a new addition to the family of multiresolution methods. In 2001, Kakarala and Ogunbona [18] generalized the Singular Value Decomposition (SVD) to MRSVD. The idea is to replace the low and high pass filters in decomposition by SVD. This is done in the following way.

1. The level variable is initialized to level $l = 1$.
2. The image is subdivided in blocks of size $p \times q$.
3. A column vector of $pq \times 1$ is generated from each block.
4. A feature vector matrix R_l of size $pq \times (MN/pq)$ is formed by stacking the column vectors.
5. Zero-mean matrix $\bar{R}_l$ is produced from R_l by mean subtraction.
6. Eigen vector matrix U_l is obtained by Singular Value Decomposition of $\bar{R}_l {\bar{R}_l}^t$.
7. The transform domain representation $\widehat{R_l} = U_l^t \bar{R}_l$ is obtained.
8. First row of $\widehat{R_l}$ represents the approximate subband and the remaining rows are the detail subbands at l^{th} level.
9. l is incremented by 1 and R_l is replaced with $\widehat{R_l(1,:)}$.

10. The steps are repeated from step 2 till the required level of decomposition is reached.

The number of subbands depend on the size of block size $p \times q$. For current implementation, block size is kept 2×2 (dyadic) to get subbands equivalent to WT.

3.4 Contourlet Transform

A key feature in multidimensional signal being the directional features, Contourlet Transform (CONT) decomposes an image into several directional subband by combining the Laplacian pyramid with a directional filter at each scale [11]. One of the classical yet popular way to obtain multiresolution decomposition of a signal is to use the *Laplacian Pyramid* introduced by Burt and Adelson [5]. The laplacian pyramid decomposes the image into a hierarchy of images that consist of different frequency band of image frequencies. It generates a lowpass image and difference between the original and the prediction, which is the bandpass image. Do [12] proposed a construction for the 2-D directional filter bank that avoids modulating the image and can simply expand the decomposition tree.

Combining the laplacian pyramid and the directional filter bank, a multiscale and directional filter decomposition can be obtained. This iterative filter bank structure is called *Contourlet filter bank*. The bandpass outputs of the laplacian pyramid, when fed to the directional filter bank, produce directional information.

For example, let $a_0[n]$ be the image. Each level of laplacian pyramid decomposes its input into a coarse and a detail image. Thus, the outputs of laplacian pyramid consist of J bandpass images $b_j[n]$ where, $j = 1, 2, ..., J$ and a lowpass image $a_J[n]$. The detail images $b_j[n]$ are fed to the l_j-level directional filter bank each producing 2^{l_j} bandpass directional images $c_{j,k}^{(l_j)}[n]$ where, $k = 0, 1, 2, ..., 2^{l_j} - 1$.

Due to the cascade structuring, the multi-scale and directional stages are independent of each other and each scale can be separately decomposed into any number of

directions provided the number of directions is a power of 2. CONT has a similarity to CT in terms of its directionality and parabolic scaling nature. For implementation, the code provided in Contourlet Toolbox by Minh. Do. is used with *9-7* filter for pyramidal decomposition stage and *pkva* filter for direction decomposition stage.

Chapter 4

Stereo Matching Algorithm based on Curvelet Decomposition and Modified Adaptive Support Weights

This chapter proposes an accurate dense disparity estimation method using curvelet decomposition and modified adaptive support weights. The method has been tested on Middlebury stereo datasets and compared with the best algorithms in current literature. The chapter consists of a brief overview of the proposed method, description of modified adaptive support weights, followed by the description of the algorithm and finally, the experimental results and comparisons.

4.1 Overview

This chapter introduces a novel method for stereo-matching based on curvelet decomposition and modified adaptive support weights. The method is termed as Stereo Matching based on Curvelet Decomposition with Modified Adaptive Support Weights (Curv+MASW). The selection of multiresolution method is based on the performance of the method to extract important features for matching across different resolutions (scales). Curvelets, as already described in Section 3.2, can efficiently represent the

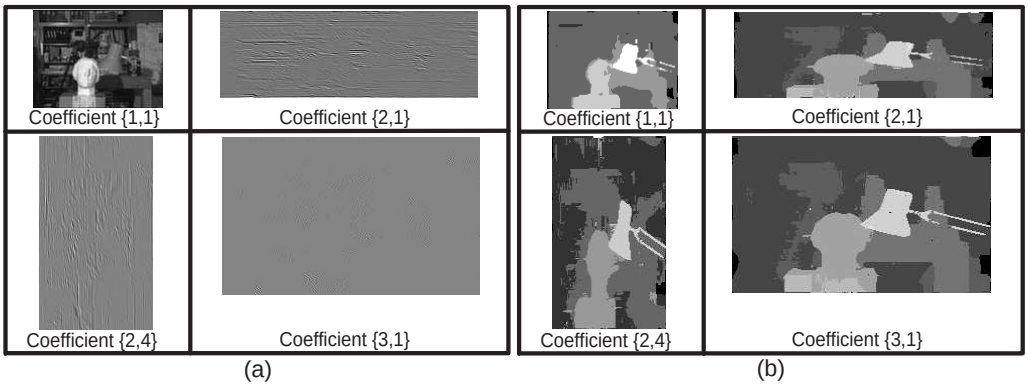

Figure 4.1: Disparity estimation in different curvelet coefficients. (a)Four curvelet coefficients are shown for left image of Tsukuba pair as gray level images. Coefficient {x,y} denotes curvelet coefficient at scale x and orientation y, (b) Their corresponding disparity maps as processed using proposed method.

curves present in the natural images. Wavelets on the other hand, can represent first order singularities (point singularities) very well but fail to represent higher order singularities (line and region) effectively. Disparity estimation is based on effective matching between the image pairs and can be considerably improved with higher number of better features to match.

Some of the coefficients as grayscale images are shown in Figure 4.1(a). There exist two advantages of using curvelets for matching. The first advantage consists of both initial disparity estimation and robust disparity estimation in multiple resolutions. The smallest curvelet coefficients, including the approximated image, provide smaller areas for searching. This results in higher search speed for an approximation of the correct location of the matched pixels. The actual location may have a small deviation, but with correct selection of the search window, the approximated disparity map is close to the ground truth. With the initial estimation of disparity, the search area in higher scales is reduced by a large amount. This reduction can increase computation speed and decrease uncertainty at the same time. The increase in speed

is evident from the method as the search area reduces. The decrease in uncertainty is due to the decrease of choices. If the choice of a matching pixel is known to be part of a complete row of the image with an unknown exact location, then there may be a number of regions in the same row with similar textures or comparable architecture, and the matching pixel may lie in any of these regions. This increases the uncertainty in finding correct matches. The process has already been shown in Figure 2.4. The search area chosen in the first resolution is broadened as the resolution increases.

The second advantage of using curvelets lies in its orientation flexibility. Figure 4.1 shows four different curvelet coefficients as images and their corresponding disparity maps as obtained using the proposed method. A close observation of the disparity maps reveals that the disparity is more prominent along the horizontal edges for coefficient {2,1}, and along the vertical edges for coefficient {2,4}. If the better matches are taken from the two disparity maps, the resulting map will have prominent disparity difference along both horizontal and vertical directions. Other subbands generally follow the same trend. This experimentally proved fact is used in the proposed method to seek optimal disparity. Also, for scale 3 there are eight orientations that yield sufficient orientation flexibility. From the figures, it is clear that the disparity shown for coefficient {3,1} resembles the original disparity to a good extent. Using higher numbers of scales and orientations does not produce significant improvement on the disparity. Thus, the proposed method is limited to third scale only.

The advantages of using curvelets make it a better choice in comparison to other multiresolution methods like the pyramid method and wavelet decomposition. The advantages also come with a burden of traversing a number of subbands for a reasonably good disparity map. In the case of wavelets, the number of subbands is limited to four for a single level and the height/width ratio remains equal. The same is not true for curvelets. But, this problem is well compensated with the improvement of quality.

4.2 Modified Adaptive Support Weights

While computing the correspondences between two pixels in left and right images, the support from neighboring pixels is used for similarity measurements. The Gestalt principles of similarity and proximity are used for computation of support weights. The adaptive support weight measurements have been proposed in Ref. [44]. The support weight is written as

$$w(p,q) = k.f_s(\Delta c_{pq}) f_p(\Delta g_{pq}), \tag{4.1}$$

where Δc_{pq} and Δg_{pq} represent the color difference and spatial distance between pixels p and q. k is a proportionality constant. $f_s()$ and $f_p()$ represent the strength of grouping by similarity and proximity, respectively. The color space is chosen to be CIELab because of the ease of 3-D representation of color. In the color space, Δc_{pq} is expressed as

$$\Delta c_{pq} = [(L_p - L_q)^2 + (a_p - a_q)^2 + (b_p - b_q)^2]^{1/2}, \tag{4.2}$$

where (L_p, a_p, b_p) are the three color components of pixel p. The strength of grouping by similarity is expressed as

$$f_s(\Delta c_{pq}) = \exp\left(-\frac{\Delta c_{pq}}{\gamma_c}\right), \tag{4.3}$$

where γ_c is a constant. Similarly, the strength of grouping by proximity has the expression

$$f_p(\Delta g_{pq}) = \exp\left(-\frac{\Delta g_{pq}}{\gamma_p}\right), \tag{4.4}$$

where Δg_{pq} can be computed using a simple euclidean distance measure between $p(x_p, y_p)$ and $q(x_q, y_q)$ as follows.

$$\Delta g_{pq} = [(x_p - x_q)^2 + (y_p - y_q)^2]^{1/2}. \tag{4.5}$$

By combining Eqs. (4.3) and (4.4), Eq. (4.1) becomes

$$w(p,q) = k \exp\left[\left(-\left(\frac{\Delta c_{pq}}{\gamma_c} + \frac{\Delta g_{pq}}{\gamma_p}\right)\right)\right]. \tag{4.6}$$

In our implementation, the curvelets basically provide a group of 2-D matrices that are treated as gray-level images. Thus, color information cannot be used to compute Δc_{pq} as in Eq. (4.2). As a replacement, the support weights (MASW) are modified by including the gray-level difference value

$$\Delta c_{pq} =\mid m_p - m_q \mid, \tag{4.7}$$

where m_p represents the gray-level value of pixel p. Using only gray-level information, the accuracy decreases, but it is balanced by the accuracy improvement due to the search operation in multiple resolution and using curvelet coefficients. For original images, the color values can be used. But, to make the process similar in each resolution, the norm of the CIELab color data is used

$$m_p = \left({L_p}^2 + {a_p}^2 + {b_p}^2\right)^{1/2}. \tag{4.8}$$

This gives a reasonably good approximation of the color data and is equivalent to the gray-level values.

The correspondence matching part is similar to any local method using cost function minimization. The support weights are computed in support windows around the pixel under consideration and its corresponding pixel in the other image. Then, the dissimilarity of the two pixels is computed by aggregating a raw matching cost with the support weights. Dissimilarity between a pixel p in the left image and a pixel p' in the right image $D(p, p')$ is expressed as

$$D(p,p') = \frac{\sum_{q \in N_p, q' \in N_{p'}} w(p,q) w(p',q') e(q,q')}{\sum_{q \in N_p, q' \in N_{p'}} w(p,q) w(p',q')} \tag{4.9}$$

where N_p and $N_{p'}$ represent support windows used for computing the support weights that belong to p and p', respectively, and, $e(p, q)$ represents the raw matching cost between pixel p and q. For the experiment, this raw matching cost has been taken as the SAD of gray values. For Eq. 4.9, $w(p, q)$ represents the support weight between pixel p and its neighbor q inside the support window of N_p, as described in Eq. 4.6.

The final disparity of the pixel p corresponds to the pixel p' for which $D(p, p')$ is minimized,

$$d_{final} = \arg\min_{p'} D(p, p') \tag{4.10}$$

4.3 Algorithm

The implementation steps in this section describe the disparity estimation using Curv+MASW. The steps are as follows.

1. Perform curvelet transform of the grayscale stereo pairs.
2. Compute the correspondence match in the lowest approximate scale of curvelets by Eq. (4.7) and (4.6).
3. Shift to the next scale, first orientation.
4. Compute scale_factor = (size of current scale image)/(size of previous scale image).
5. For each pixel, divide the coordinate of the pixel by scale_factor to get the coordinates in the previous scale. Then, compute the initial_curr_disparity = scale_factor * disparity in the previous scale.
6. Take a range of search around the initial_curr_disparity.
7. Find the best match using MASW from Eq. (4.6).
8. If this is the last orientation, go to step 10.
9. Shift to the next orientation. Go to step 5.
10. For each pixel, there is a match in each orientation at the current scale. Best match can be found by a normal correlation value check or a left-to-right and right-to-left disparity consistency check.

11. If this is the last curvelet scale, go to step 12, else go to step 3.

12. compute final disparity map using MASW for the original image.

Here, a brief description of step 10 is necessary to describe the left-to-right and right-to-left disparity consistency check. For the process, the left (reference) image disparity map is computed by matching with right (target) image. Next, the reference and target images are interchanged and disparity is again computed for right (new reference) image by matching with left (new target) image. Then, the two disparity maps are checked against each other to examine whether, a disparity is consistent for a pair of image points in both the disparity maps. Any inconsistency corresponds to a ill-matched pixel pair and can be removed.

The implementation does not depend on any initial disparity estimation provided externally. The initial estimation is already performed in the lowermost scale, i.e., the approximation. The approximated disparity map is then improved in each scale using different orientations of curvelets.

The initial disparity map is not totally correct due to the approximated scale but can define the range of search, and also has the limit of maximum search. With the curvelet coefficients obtained at scale = 3, orientation = 8, the approximated image for Tsukuba pair has a dimension of 97×129. In this dimension, using the window size 21×21, the maximum disparity obtained is 5. If this is multiplied by a scale factor equal to the ratio of the original image size and approximated image size, the disparity in the original image size has a maximum range of 15. This reduces the search area at the border regions, even if the approximate disparity map does not have any disparity value in these areas, as shown in Figure 4.2(a) and 4.2(c).

The support weights reduce the fattening effect by a considerable amount. Comparisons have been done for the results of the initial disparity maps with support weights [as shown in Figure 4.2(a) and 4.2(c)] and with NCC [shown in Figure 4.2(b) and 4.2(d)]. The Cones and Aloe disparity maps for the approximated image created

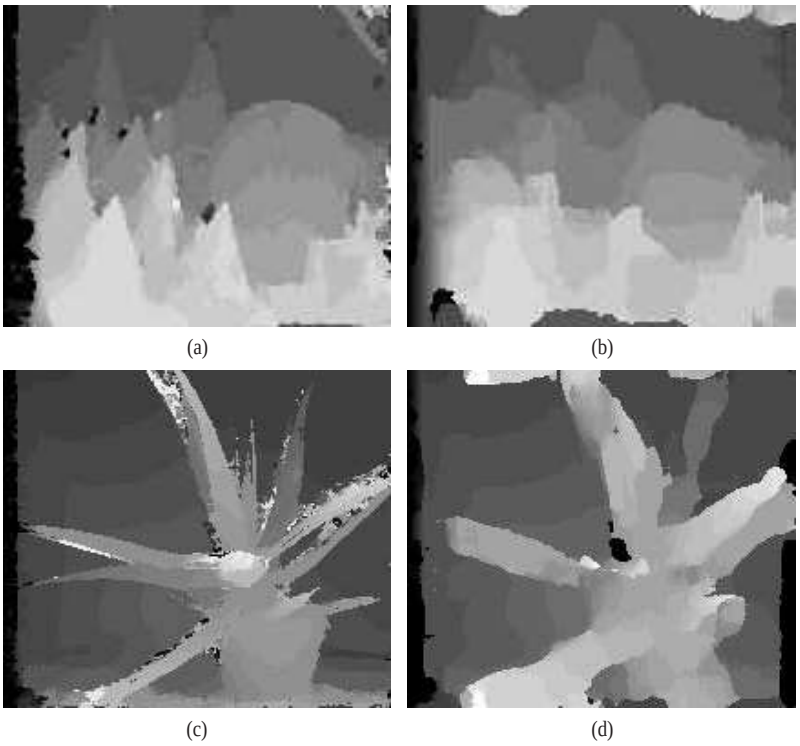

Figure 4.2: Comparison of initial disparity map generated with support weight and NCC. Figure (a) Cones - with support weights, (b) Cones - with normalized cross correlation. Similarly, (c) Aloe - with support weights and (d) Aloe - with normalized cross correlation.

with support weight are clearer and have sharp disparity changes along the edges. The disparity maps with NCC show poor results along the edges and occluding areas. The reason for the reduction of the fattening effect is well explained in the paper by Yoon and Kweon [44].

The discontinuities in the images at the occlusion boundaries are not properly recovered in the disparity map using local methods due to inaccurate support window size selection. Using support weights, this limitation is largely removed. The support weights are based on the color similarity and distance information of the neighbouring pixels and they can well predict the disparity in the discontinuities, even in a fixed

window size. It is further improved with the use of unique curvelet coefficients, which add the edge and orientation information to the pixels. Thus, by combining the methods, the disparity maps have been highly improved and are comparable to the results of global methods.

4.4 Experimental Results

The experiments have been carried out on the Middlebury 2001, 2003, 2005 and 2006 datasets [16, 30, 31, 32]. The comparison has been divided into three different parts.

1. A comparison has been done with the state of the arts methods in literature for the four datasets in the Middlebury website namely Tsukuba, Venus, Teddy and Cones. These four image pairs have been used by all other algorithms as a common platform to compare the performances. Thus, these pairs have been obvious choices for the comparison.

2. A comparison with Adaptive Support Weights (adaptWeight) has been done to show the relative improvement due to the curvelet decomposition.

3. Finally, four of the complex image pairs namely, Teddy, Cones, Sawtooth and Map have been chosen for a comparison with stereo-matching based on Wavelet with Modified Adaptive Support Weights (Wave+MASW) to show the relative improvement of curvelets over wavelets.

4.4.1 Comparison with State of the arts Methods

In this section, the experimental results and comparisons are presented for the four image pairs from Middlebury 2001 and 2003 datasets. The parameters in the tests are specified as follows: Size of support window = 33×33, $\gamma_p = 36$, $\gamma_c = 7$ and $k = 1.5$. For the smallest approximation, window size has been taken as 21×21. The search

area for higher scales has ±10 pixels deviation from the original disparity position of the initial disparity map and reduces to ±5 pixels for the original size image pairs.

The experimental results are shown in Figure 4.3. The comparison results with other methods are tabulated in Table 4.1. The numbers in the table represent the percentage of bad pixels compared to the groundtruth: for all pixels "all," pixels in the non-occluded regions "nonocc," and pixels near depth discontinuities "disc." These results have been evaluated with the formula for percentage of bad pixels defined as

$$\frac{1}{N}\sum_{x,y}(\mid d_c(x,y) - d_t(x,y) \mid) > \delta_{thresh},$$

where N is the total number of pixels in the image, d_c and d_t represent the computed and groundtruth disparity maps, respectively, and δ_{thresh} is the threshold for bad pixels (usually equal to 1.0). Thus, Lower number indicates higher accuracy. As can be seen from the table, proposed method outperforms most of the local methods for stereo-matching. It it also better than some of the global-based methods like RegionTreeDP, RealtimeBP, GC+occ, etc. Of course there are still some local and global methods better than Curv+MASW. That compels for more improvements that are presented in Chapter 5.

4.4.2 Comparison with Adaptive Support Weights

In this section, the main concern is to show the improvement using curvelets. Thus, more results have been provided with other datasets from the Middlebury 2005 and 2006 database in Figure 4.4, along with the results using the adaptWeight. The results are shown column wise with each column representing the left image of a stereo pair, its disparity map generated with Curv+MASW, adaptWeight, and the corresponding groundtruth. The estimated percentage of bad pixels are tabulated in Table 4.2. From the table, it is evident that Curv+MASW has lower amount of error.

From the comparison result in Table 4.1, it is clear that Curv+MASW has better performance than the adaptWeight, RegionTreeDP,RealtimeBP, etc. methods. In

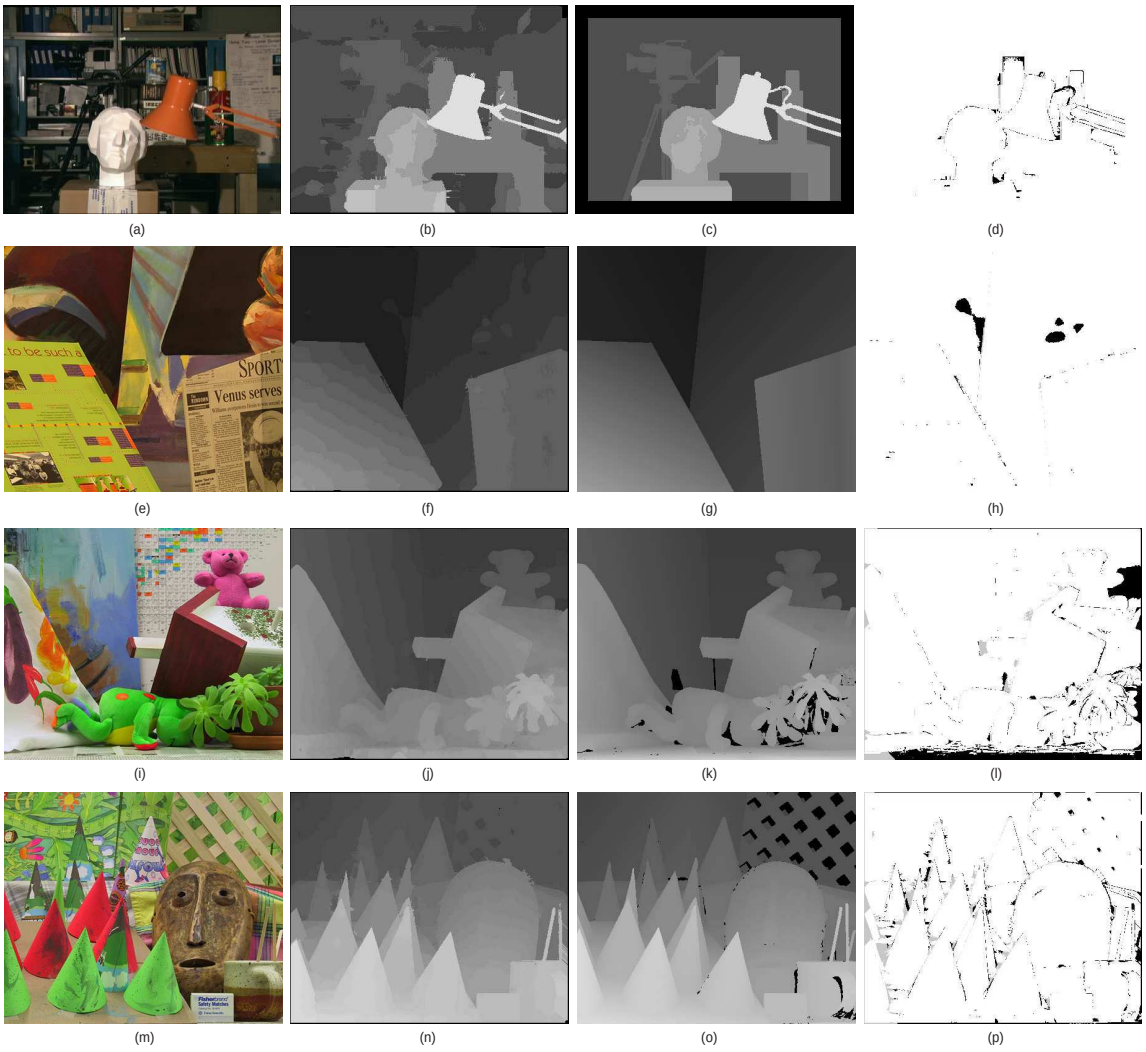

Figure 4.3: Dense disparity maps for the Middlebury images using Curv+MASW and their corresponding ground truths. (a) Tsukuba - left image, (b) Corresponding Curv+MASW, (b) Ground truth, (d) Bad pixels, (e) Venus - left image, (f) Corresponding Curv+MASW, (g) Ground truth, (h) Bad pixels, (i) Teddy - left image, (j) Corresponding Curv+MASW, (k) Ground truth, (l) Bad pixels, (m) Cones - left image, (n) Corresponding Curv+MASW, (o) Ground truth, (p) Bad pixels.

Table 4.1: Comparison with state of the arts methods, the proposed method is denoted as Curv+MASW

Methods \ Images	Tsukuba			Venus			Teddy			Cones		
	nonocc	all	disc	nonocc	all	disc	nonocc	all	disc	nonocc	all	disc
CoopRegion [41]	0.87	1.16	4.61	0.11	0.21	1.54	5.16	8.31	13.0	2.79	7.18	8.01
DoubleBP [42]	0.88	1.29	4.76	0.13	0.45	1.87	3.53	8.30	9.63	2.90	8.78	7.79
GeoSup [17]	1.45	1.83	7.71	0.14	0.26	1.90	6.88	13.2	16.1	2.94	8.89	8.32
symBP+occ [36]	0.97	1.75	5.09	0.16	0.33	2.19	6.47	10.7	17.0	4.79	10.7	10.9
AdaptDispCalib [15]	1.19	1.42	6.15	0.23	0.34	2.50	7.80	13.6	17.3	3.62	9.33	9.72
Curv+MASW	**1.40**	**1.84**	**7.42**	**1.00**	**1.11**	**4.42**	**7.85**	**8.84**	**16.8**	**3.82**	**6.22**	**8.24**
CostAggr+occ [28]	1.38	1.96	7.14	0.44	1.13	4.87	6.80	11.9	17.3	3.60	8.57	9.36
RegionTreeDP [24]	1.39	1.64	6.85	0.22	0.57	1.93	7.42	11.9	16.8	6.31	11.9	11.8
adaptWeight [44]	1.38	1.85	6.90	0.71	1.19	6.13	7.88	13.3	18.6	3.97	9.79	8.26
SegTreeDP [10]	2.21	2.76	10.3	0.46	0.60	2.44	9.58	15.2	18.4	3.23	7.86	8.83
RealtimeBP [43]	1.49	3.40	7.87	0.77	1.90	9.00	8.72	13.2	17.2	4.61	11.6	12.4
GC+occ [23]	1.19	2.01	6.24	1.64	2.19	6.75	11.2	17.4	19.8	5.36	12.4	13.0

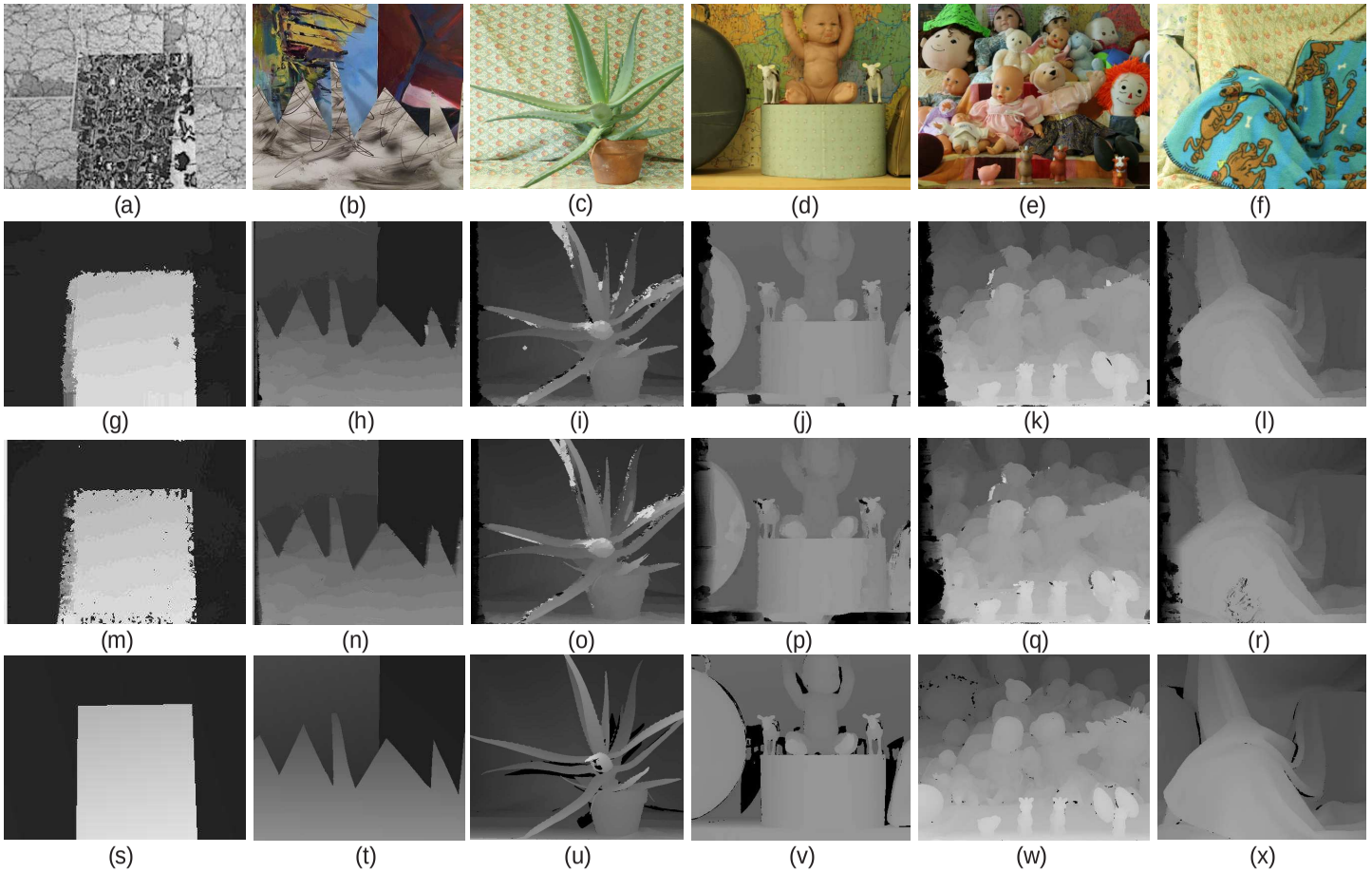

Figure 4.4: Comparison of Curv+MASW with adaptWeight for Middlebury datasets. Each column consists of a different set of left image of a stereo pair, disparity map produced by Curv+MASW, disparity map produced by adaptWeight and the corresponding ground truth. Column 1 to 6 represent the figures for Map, Sawtooth, Aloe, Baby, Dolls and Cloth pairs respectively.

Table 4.2: Comparisons of Curv+MASW with adaptWeight

Images / Methods	Map	Sawtooth	Aloe	Baby	Dolls	Cloth
	% of bad pixels					
Curv+MASW	0.8212	0.5382	0.1666	0.2692	0.3595	0.2103
adaptWeight	0.8554	0.5472	0.2114	0.3960	0.4494	0.2926

comparison to adaptWeight, the Tsukuba and Venus pairs do not show considerable improvements, while the Cones and Teddy pairs are significantly improved. The reason for this lies in the initial disparity map creation. For Tsukuba and Venus, the maximum disparity is much smaller compared to Cones and Teddy. For smaller disparities, the initial disparity map creation process does not produce effectively better results than the other methods. Nevertheless, with increasing amount of disparity, the method performs better than others due to the uniqueness of the curvelet coefficients in lower resolutions. This difference is prominent in Cones and Teddy. The Cones pair has been listed in the top two in the Middlebury evaluation for all matching regions. The time complexity does not change too much with the variation of image size, because the main searching operation is performed in lowest scales, and higher scale operations have fixed search areas. The time complexity can be reduced greatly if some of the higher order scales and orientations are omitted, e.g., for third scales, some orientations have repetitive textures and can be omitted with a small decrease in accuracy. An experiment on time complexity analysis of different multiresolution based correspondence methods have been performed and experimental results are shown in Chapter 6. The results shown before can be further improved by using a higher number of scales and angles or using better outlier removal methods in combination. The simple combination of curvelets and support weights is fast and can produce disparity maps suitable for applications that do not need highly accurate disparity maps. An example would be a preprocessing step for a global method.

Table 4.3: Comparison of Curv+MASW and Wave+MASW for Middlebury datasets

Images / Methods	Teddy	Cones	Map	Sawtooth
	% of bad pixels			
Curv+MASW	0.3523	0.3357	0.8212	0.5382
Wave+MASW	0.3824	0.3802	0.9711	0.5442

4.4.3 Comparison with Wavelets

An extra comparison has been included in the research to visualize the improvement in performance with the inclusion of curvelets over the inclusion of wavelets. The results are shown for Teddy, Cones, Map and Sawtooth image pairs because the image pairs include the following properties - left and right half occlusions, sharp structures (scope for fattening effect), repetitive and non-textured regions. The results are shown in Figure 4.5 and quantitative analysis in terms of percentage of bad pixels matched, is shown in Table 4.3. As it was claimed before, the results for Curv+MASW are better than those of Wave+MASW.

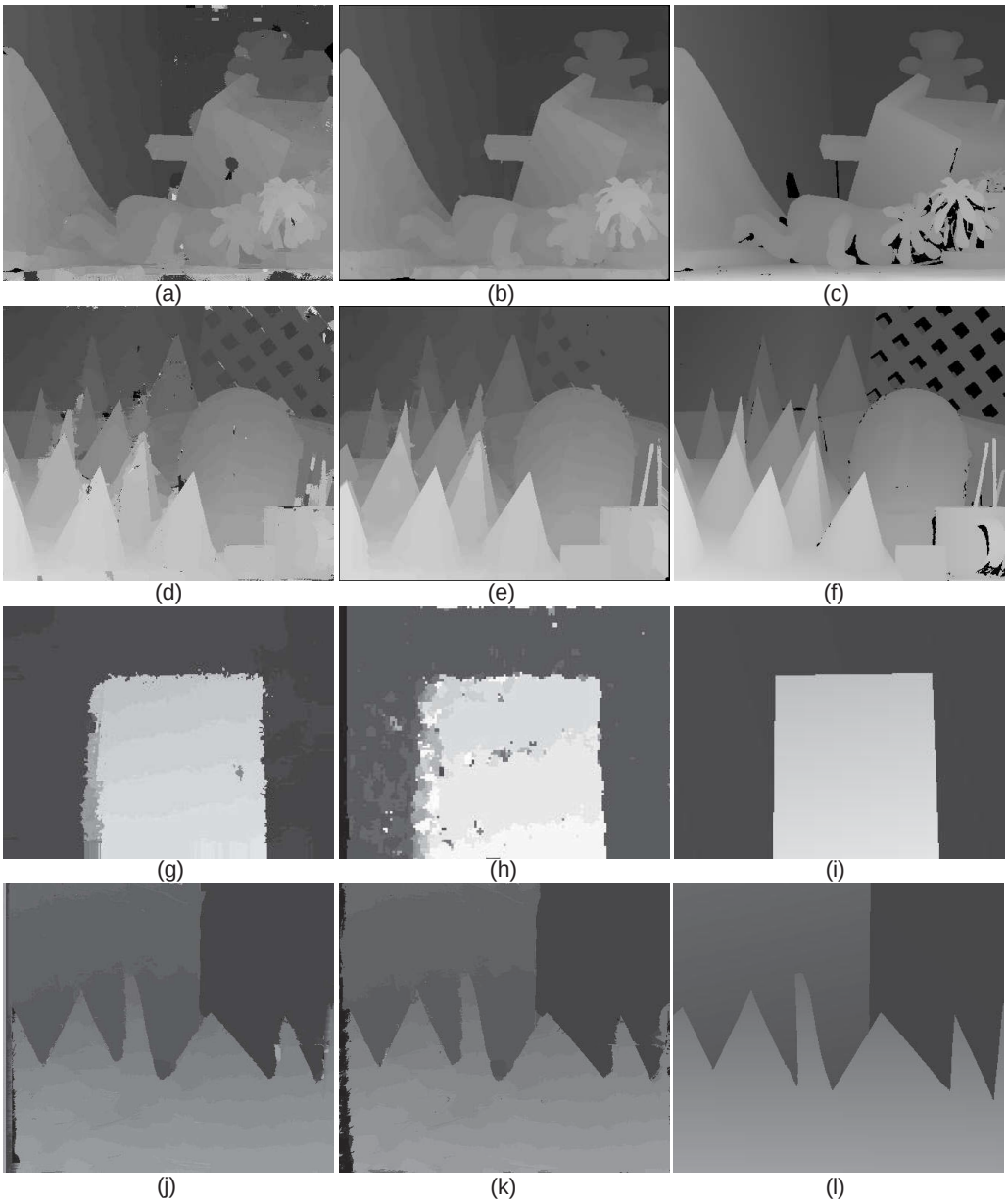

Figure 4.5: Disparity maps for Teddy, Cones, Sawtooth and Map pairs produced by Curv+MASW and Wave+MASW. Each row shows the result of Wave+MASW followed by the result of Curv+MASW and the corresponding ground truth.

Chapter 5

Stereo Correspondence based on Curvelet Decomposition, Support Weights and Disparity Calibration

The proposed method Curv+MASW is better than most of the local methods but still not comparable to top global methods. This fact acted as a motivation to improve the method. Curv+MASW does not take advantages of the continuity constraint discussed in Section 2.2 which states that disparity smoothly varies on object surfaces and discontinuities occur at the object boundaries. The results of Curv+MASW, if carefully observed, are poor at the object boundaries and continuous surfaces with narrow edges. These parts can be improved considering the continuity constraint.

The current section proposes the use of Disparity Calibration (DC) with the process of stereo-matching based on Curvelet Decomposition with Modified Adaptive Support Weights (Curv+MASW). The overview of the method is given in Section 5.1.DC is based on disparity voting that effectively uses the continuity constraint and is explained in the Section 5.2. This is followed by the description of the algorithm in Section 5.3. Finally the experimental results to support the claims are provided in Section 5.4.

5.1 Overview

This chapter proposes a novel method for highly accurate correspondence estimation termed Stereo Matching based on Curvelet Decomposition with Modified Adaptive Support Weights and Disparity Calibration (curv+MASW+DC). The method is an extension of Curv+MASW and merges the use of continuity constraint to improve the results.

5.2 Disparity Calibration

Disparity calibration is based on the assumption that points with similar color or short spatial distance have similar disparity. For any point in the image, a calibration window is chosen around it. The size of the window is optimized according to the criteria that the pixels in the window have similar color or short distance. Then, the occurrence number or frequency of every possible disparity is checked in the window. The disparity with the highest occurrence number is assigned to the point. The process of disparity calibration can be summarized as follows:

1. Find the maximum and minimum disparities $(d_{\max}, d_{\min})$.

2. For each disparity d in $[d_{\max}, d_{\min}]$, define zero matrix C_d with the same size as the image. There will be a series of matrices.

3. In the original disparity map D, find the pixels with disparity d, assign 1 to the corresponding elements in C_d.

4. Result disparity $D(i,j)$ for pixel (i,j) is expressed as

$$\begin{aligned} D(i,j) &= \arg\max_d \left\{ \sum_{x=-\frac{1}{2}(W_x-1)}^{x=\frac{1}{2}(W_x-1)} \sum_{y=-\frac{1}{2}(W_y-1)}^{y=\frac{1}{2}(W_y-1)} \right. \\ &\left. \times w((i,j),(i+x,j+y)) \times C_d(i+x,j+y) \right\} \end{aligned} \tag{5.1}$$

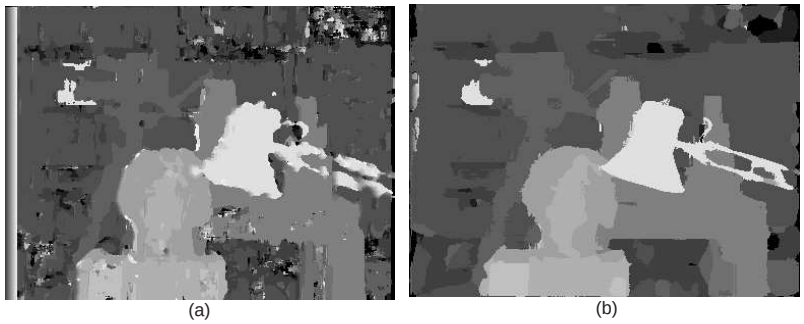

Figure 5.1: Improvement with application of DC.(a) Tsukuba disparity map using SAD, (b) Tsukuba disparity map using SAD+DC.

where W_x and W_y represent the width and height of the calibration window, respectively, and $w(x,y) = \exp\left[-\left(\frac{\Delta c_{pq}}{\gamma_i} + \frac{\Delta g_{pq}}{\gamma_p}\right)\right]$ represent the Modified Adaptive Support Weights (MASW) of pixel (x,y) with $\gamma_i < \gamma_c$ to make the color constraint stricter.

The disparity range $(d_{\max}, d_{\min})$ is already estimated from the initial and final disparity calibration. Thus, the implementation of disparity calibration is straightforward.

The main contribution of DC to the algorithm lies in the voting principle of disparity assignment in step 4 that consists of multiplication of the cost function with C_d. This implicitly uses the continuity constraint because, a voting based on neighboring pixels' disparity assignment implies the use of their dependencies. An improvement using DC is shown in Figure 5.1. The increase in accuracy is clear from the figures.

5.3 Algorithm

The implementation steps of curv+MASW+DC are quite similar to that of Curv+MASW and described as follows.

1. Perform curvelet transform of the grayscale stereo pairs.

2. Compute the correspondence match in the lowest approximate scale of curvelets by Eq. (4.9).

3. Shift to the next scale, first orientation.

4. Compute scale_factor = (size of current scale image)/(size of previous scale image).

5. For each pixel, divide the coordinate of the pixel by scale_factor to get the coordinates in the previous scale. Then, compute the initial_curr_disparity = scale_factor * disparity in the previous scale.

6. Take a range of search around the initial_curr_disparity.

7. Find the best match using MASW from Eq. (4.9).

8. If this is the last orientation, go to step 10.

9. Shift to the next orientation and go to step 5.

10. For each pixel, there is a match in each orientation at the current scale. Best match can be found by a normal correlation value check or a left-to-right and right-to-left disparity consistency check.

11. Smooth the initial_curr_disparity using disparity calibration.

12. If this is the last curvelet scale, go to step 13, else go to step 3.

13. For original image pair, use the norm of CIELab color images using Eq. (4.8) and Eq. (4.9) and compute the final disparity map using MASW.

    Compared to the algorithm in Section 4.3, step 11 is the addition. After the enhanced matching in curvelet subbands using modified support weights, the disparity map is further smoothed by DC by assigning disparity to the pixels with wrong disparity or without disparity.

5.4 Experimental Results

Curv+MASW in combination with DC has been carried out on the Middlebury dataset, and results have been compared with other methods. The results are shown in Figure 5.2 and comparisons with some of the best global and local methods are tabulated in Table 5.1, where curv+MASW+DC stands for the proposed method, and Local Stereo Matching with Adaptive Support-Weight, Rank Transform and Disparity Calibration (AdaptDispCalib) denotes the original Disparity Calibration. Currently, curv+MASW+DC performs best among the local area based methods.

As seen from the results, curv+MASW+DC significantly improved the Cones and Teddy disparity maps. The improvement is due to the early disparity estimation at the lower resolutions, the uniqueness of the curvelet coefficients, and smoothing of the disparity map using DC. For AdaptDispCalib, the disparity range has to be provided for a smaller search range and reduction of disparity ambiguity. For curv+MASW+DC, this range is readily provided by lower resolution data. The results are better compared to all of the existing local methods proving to be comparable to top global methods.

The results also confirm the previous conclusion that the proposed algorithm performs better with an increased amount of disparity. The improvements on Cones and Teddy images scored the highest rank in the Middlebury evaluation for all region matching.

curv+MASW+DC itself produces highly accurate disparity maps and involves a relatively lower number of parameters dependent on image size and type, in comparison to other methods. Thus, the proposed method can be thought of as a complete stereo-matching process, unlike Curv+MASW.

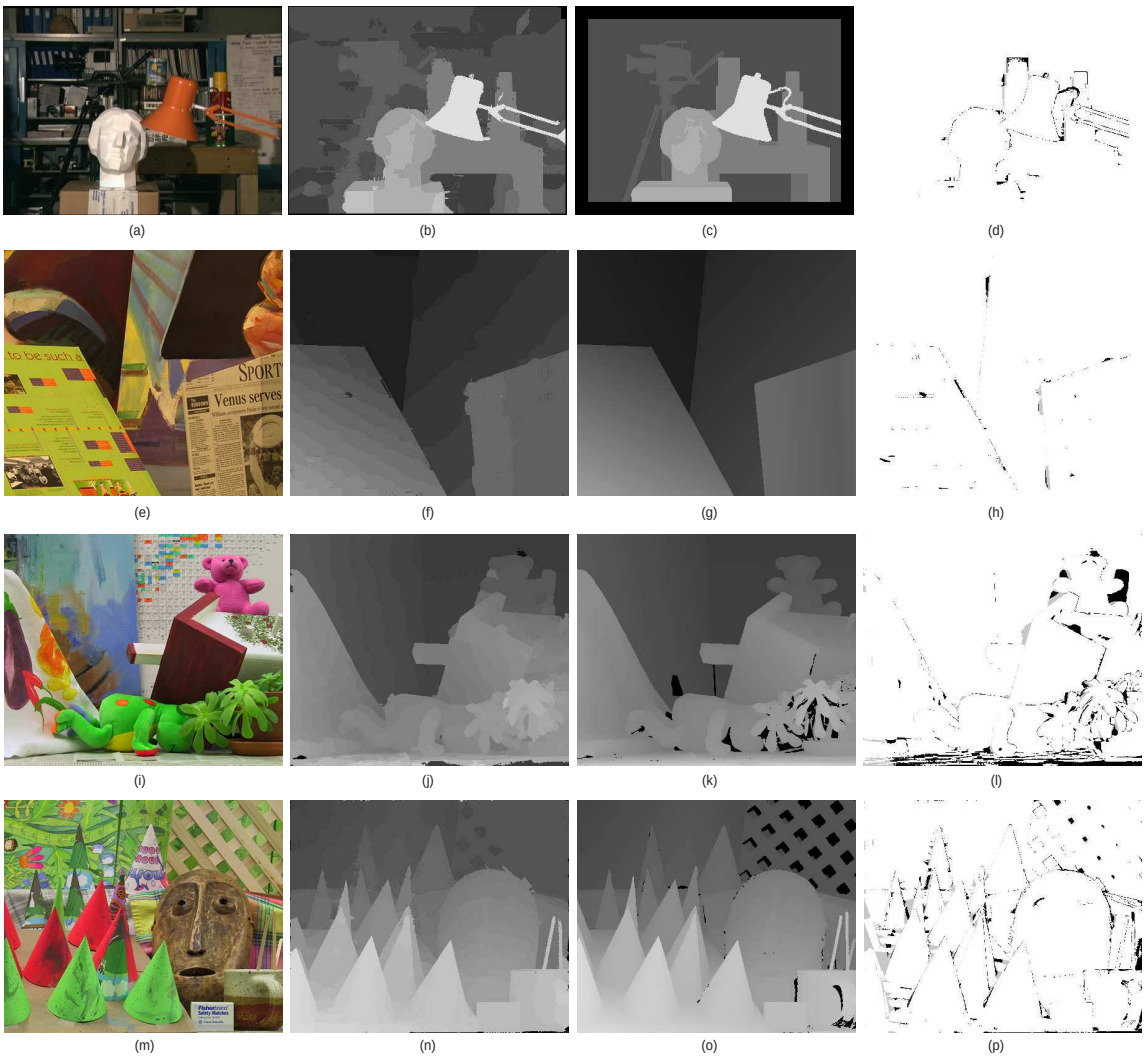

Figure 5.2: Dense disparity maps for the Middlebury images using curv+MASW+DC, the corresponding ground truths and the bad pixels matched. (a) Tsukuba - left image, (b) Corresponding curv+MASW+DC, (b) Ground truth, (d) Bad pixels, (e) Venus - left image, (f) Corresponding curv+MASW+DC, (g) Ground truth, (h) Bad pixels, (i) Teddy - left image, (j) Corresponding curv+MASW+DC, (k) Ground truth, (l) Bad pixels, (m) Cones - left image, (n) Corresponding curv+MASW+DC, (o) Ground truth, (p) Bad pixels.

Table 5.1: Comparison with state of the arts methods, the proposed method is denoted as curv+MASW+DC

Methods \ Images	Tsukuba			Venus			Teddy			Cones		
	nonocc	all	disc	nonocc	all	disc	nonocc	all	disc	nonocc	all	disc
CoopRegion [41]	0.87	1.16	4.61	0.11	0.21	1.54	5.16	8.31	13.0	2.79	7.18	8.01
DoubleBP [42]	0.88	1.29	4.76	0.13	0.45	1.87	3.53	8.30	9.63	2.90	8.78	7.79
curv+MASW+DC	**1.36**	**1.80**	**7.18**	**0.46**	**0.85**	**4.17**	**5.65**	**6.71**	**14.8**	**3.23**	**5.51**	**8.45**
GeoSup [17]	1.45	1.83	7.71	0.14	0.26	1.90	6.88	13.2	16.1	2.94	8.89	8.32
symBP+occ [36]	0.97	1.75	5.09	0.16	0.33	2.19	6.47	10.7	17.0	4.79	10.7	10.9
AdaptDispCalib [15]	1.19	1.42	6.15	0.23	0.34	2.50	7.80	13.6	17.3	3.62	9.33	9.72
Curv+MASW	**1.40**	**1.84**	**7.42**	**1.00**	**1.11**	**4.42**	**7.85**	**8.84**	**16.8**	**3.82**	**6.22**	**8.24**
CostAggr+occ [28]	1.38	1.96	7.14	0.44	1.13	4.87	6.80	11.9	17.3	3.60	8.57	9.36
RegionTreeDP [24]	1.39	1.64	6.85	0.22	0.57	1.93	7.42	11.9	16.8	6.31	11.9	11.8
adaptWeight [44]	1.38	1.85	6.90	0.71	1.19	6.13	7.88	13.3	18.6	3.97	9.79	8.26

Chapter 6

A Comprehensive Comparison Framework

A brief discussion of multiresolution methods and their application is stereo-matching has already been discussed in Chapter 2 and 3. Studies reveal that the multiresolution based stereo-matching methods existing in literature are relatively independent. Presently, they do not relate to a continuous progress in the research. Relatively less research has been carried out to compare different multiresolution methods and how detail subbands contribute to enhance the performance of the matching process. As a result, the relative advantages and disadvantages of a particular multiresolution method in disparity estimation are quite incomprehensible.

The goal of the present work is to provide a comparative study of different multiresolution methods for disparity estimation to highlight their expediency and suitability for the process and their relative performances. The primary reason for such a study is to highlight the progress in a relatively less popular domain and motivate the researchers to make more advancements. Also, it allows to analyze each algorithm and improve their performances. Finally, it highlights the limitations of a method with respect to the others and facilitate the choice of an appropriate multiresolution method for specific needs.

This chapter is divided into three sections. Section 6.1 gives a brief overview of the work followed by the description of the proposed comparison framework in

Section 6.2. Finally, Section 6.3 provides the qualitative and quantitative comparison results.

6.1 Overview

A comparison framework has been proposed in this chapter to evaluate the performances of different multiresolution methods. The framework is divided into three parts viz. fixWin, segWin and segWinApprox. To the best of our knowledge, this work is the first attempt to explore and evaluate the performances of different multiresolution methods in disparity estimation by the above mentioned framework.

For the study, four different multiresolution methods are chosen - Daubechies Wavelet Transform (WT) from the wavelet family, Curvelet Transform (CT), Multiresolution Form of Singular Value Decomposition (MRSVD) and Contourlet Transform (CONT). The comparison has been done on well-known Middlebury stereo datasets.

6.2 Proposed Framework

In order to evaluate performances, three different frameworks are used for a number of reasons. A comparison among methods does not only rely on the performance of the methods but also on the comparison framework used. Using three different frameworks, a better knowledge on the performance of the methods can be obtained. Also, the results in Section 6.3 show that by using better cost functions, the results can be greatly improved. In Section 6.2.1, the support window is kept as fixed size and include all subbands of multiresolution method. Next, segmented window based cost function is used for more accurate disparity estimation in Section 6.2.2. Finally, in Section 6.2.3, the classical approximate band based disparity estimation is used to highlight the effects of simple multiple resolution searching. In all the cases, SAD is used as the cost function. Further, in order to keep an equivalent level of WT and

MRSVD decomposition, single level is used in Section 6.2.1 and 6.2.2 whereas 2-level is used in sub-section 6.2.3. As per as CT and CONT are concerned, one cannot use level of decomposition because they are decomposed into scales and orientations. Hence, scale 2 and orientation 8 are used for Section 6.2.1 and 6.2.2 whereas scale 3 and orientation 8 are used for Section 6.2.3.

6.2.1 Fixed Window SAD

Fixed Window based Sum of Absolute Differences (fixWin) uses SAD as cost function and refines disparity in multiple subbands over a fixed size window in the following algorithm:

1. Perform multiresolution analysis of the grayscale stereo pairs.
2. Compute the correspondence match in the approximate image using SAD.
3. Shift to the subband.
4. Compute scale_factor = (size of current subband image)/(size of previous subband image).
5. For each pixel, divide the coordinate of the pixel by scale_factor to get the coordinates in previous scale. Then, compute the initial_curr_disparity = scale_factor * disparity in previous scale.
6. Take a range of search around the initial_curr_disparity.
7. Find the best match using SAD.
8. If this is the last subband, go to step 10.
9. Shift to the next subband. Go to step 5.

10. Take a range of search in final image based on the refined disparity map recovered after step 9.

11. Find the final disparity map using SAD.

For WT and MRSVD, only four subbands are obtained while CT and CONT yield a number of subbands in different scales and orientations. In the algorithm, it is assumed that if current subband is at scale s and orientation o, then next subband will be at scale s and orientation $o+1$, if orientation $o+1$ exists, else it will be the first orientation in next scale. The algorithm implicitly mentions two parameters - the search range for refinement and the fixed support window size. To keep consistency in quality of estimated disparity, both are kept constant for each comparison.

6.2.2 Segmented Window SAD

The major disadvantage of fixed window is the fattening effect in depth discontinuities which has already been mentioned in Section 4.3. As all points in a support window are considered to be of similar depth, the regions with discontinuity in depth are poorly matched. If the pixels in the window are kept to same region, the results can be improved. Based on This idea, this section describes Segmented Window based Sum of Absolute Differences (segWin) that works particularly better in these regions and can largely remove fattening effect. For this method, color images are used to produce multiresolution output of each color band separately. Each color band yields a set of multiresolution subbands. In another way, each subband has three gray value components each corresponding to a color. These *tricolor* subbands are used to segment the image into different regions by simple thresholding. The pseudo code below shows the procedure for the same.

1. $SAD = 0$;

2. For i IN $-W$ TO W

3. For j IN $-W$ TO W

4. IF dist$(sBL(x+i, y+j, :), sbL(x, y, :)) < thresh$

5. $SAD = SAD + (|sBL(x+i, y+j, :) - sBR(x+i+d, y+j, :)|)$

6. END IF

7. END FOR

8. END FOR

Where $sBL(\cdot)$ and $sBR(\cdot)$ denote the *tricolor* subbands with $sBL(x, y, :)$ denoting three components of pixel (x, y), dist is a function that returns the euclidean distance between the two component vectors of two pixels, d denotes the disparity, $(2 * W + 1)$ is the support window size and $thresh$ is the threshold for segmentation. Only the pixels that meet the criteria in IF condition are chosen for SAD computation.

The extra parameter involved in this method is the threshold value. This value needs to be kept at minimum range to remove the interregional interference.

6.2.3 Approximate Band based Matching

The study is concluded with classical approximate band searching that ignores detail subbands and searches in approximate bands progressively till it reaches the original image resolution. The search region is projected into higher resolutions and searching is limited only to this region. This work is an effort to highlight the effect of low filtered approximate band for stereo-matching. The method Segmented Window based Sum of Absolute Differences on Approximate Band (segWinApprox) uses segWin for approximate band matching and the comparison has been done in single and double level of decomposition as already described before.

Table 6.1: Comparisons of methods based on fixWin

Methods / Images	WT	CT	MRSVD	CONT	OrigSAD
	% of bad pixels				
Tsukuba	10.44	9.38	10.53	10.57	11.82
Cones	21.79	20.81	20.41	21.69	21.55

6.3 Comparison Results

The results are divided into three sections based on the framework used. The sections are - Fixed Window based Results, Segmented Window based Results and Approximate Band Results. The well-known Middlebury datasets Tsukuba [31] and Cones [32] are used for generating comparative results. The window size has been kept 9×9 in original scale, 7×7 for level 1 of decomposition for fixWin and 33×33 in original scale and all subbands for segWin and segWinApprox.

6.3.1 Fixed Window based Results

In Table 6.1, the comparison of different methods is provided for fixWin. The results are also compared with Original SAD based Stereo Matching (OrigSAD) result in terms of percentage of bad pixels.

In Table 6.2, the execution times of different methods are compared for fixWin in terms of seconds. For time complexity, it is important to mention that all methods were run in a Pentium 4, 2 GHz machine with MATLAB. It is evident from the results that multiresolution methods provide comparative results in faster computations. The qualitative results can be seen from Figure 6.1.

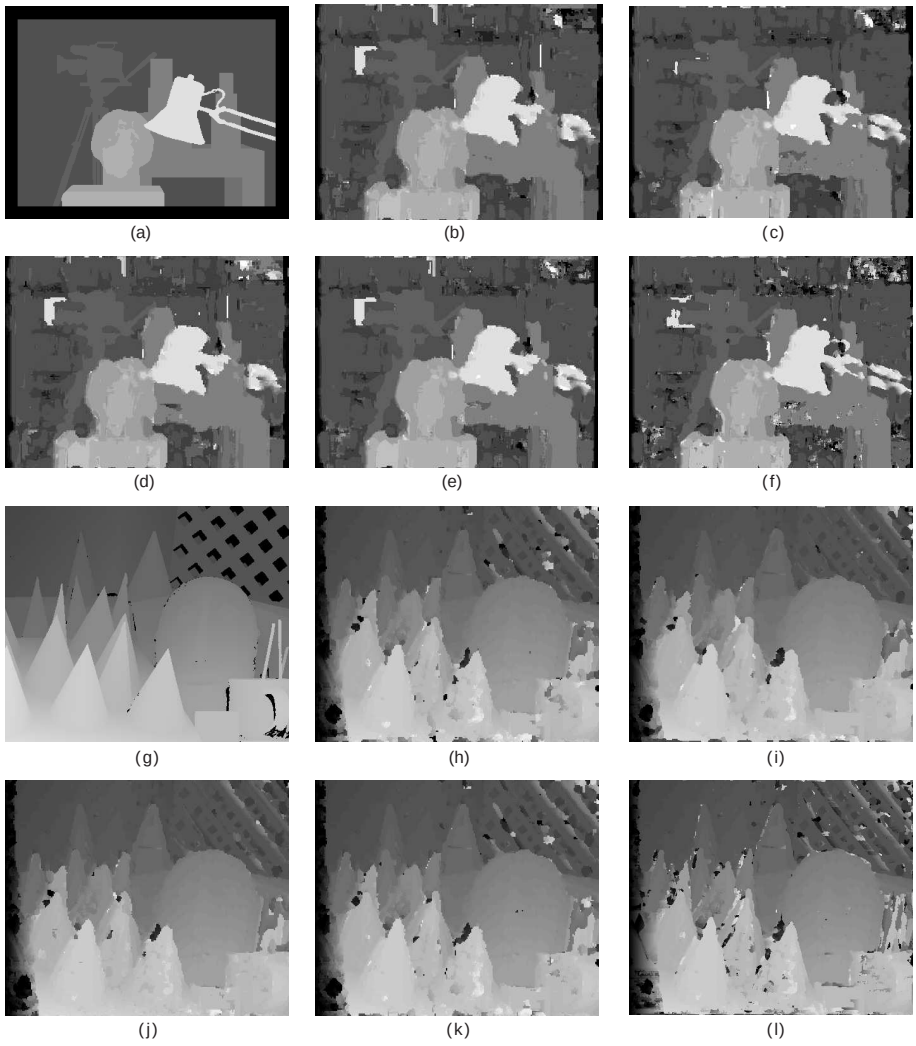

Figure 6.1: Disparity maps generated using fixWin. (a),(g) ground truth, (b),(h) WT, (c),(i) CT, (d),(j) MRSVD, (e),(k) CONT and (f),(l) OrigSAD.

Table 6.2: Comparison execution time of different methods based on fixWin

Images \ Methods	WT	CT	MRSVD	CONT	OrigSAD
Tsukuba	13.9313	19.3346	14.1931	17.3685	25.0329
Cones	28.2972	44.7665	28.5641	28.7104	130.3596

Table 6.3: Comparison of methods based on segWin

Methods / Images	WT	CT	MRSVD	CONT
	% of bad pixels			
Tsukuba	6.23	5.72	6.35	6.63
Cones	19.31	17.69	17.64	19.68

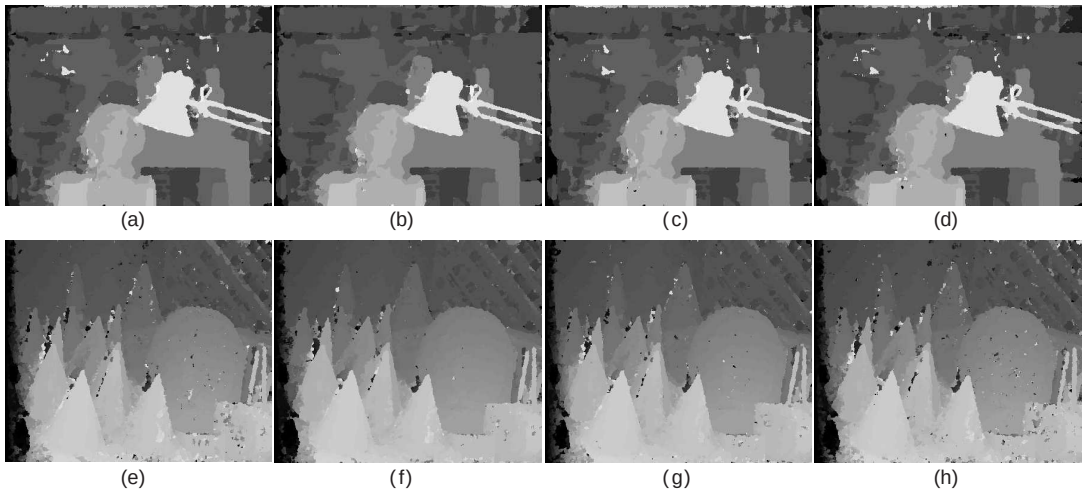

Figure 6.2: Disparity maps generated using segWin. (a),(e) WT, (b),(f) CT, (c),(g) MRSVD and (d),(h) CONT.

6.3.2 Segmented Window based Results

Table 6.3 shows the comparison among different multiresolution methods based for segWin. The results, as can be seen from Figure 6.2, are much better than fixWin and are quite immune to the fattening effect. As segWin is much slower compared to fixed window based method, time complexity analysis is not provided. Of course, the execution time depends on the platform and can be improved.

Table 6.4: Comparisons of methods based on segWinApprox

Images \ Methods	WT		CT		MRSVD		CONT	
	% of bad pixels							
	L1	L2	L1	L2	L1	L2	L1	L2
Tsukuba	8.73	9	8.42	8.33	9.2	9.4	8.78	8.85
Cones	23.59	27.27	22.72	25.4	21.24	22.61	23.51	27.54

6.3.3 Approximate Band based Results

Table 6.4 shows the quantitative comparison for segWinApprox and the qualitative results are provided in Figure 6.3. The results are not as good as segWin due to the unavailability of other subbands.

6.3.4 Analysis of Results

Comparison results reveal that CT and MRSVD generally give better results than those of CONT and WT. Moreover, the results of CT are better in lower search range as in Tsukuba pairs and MRSVD provides better results in higher search ranges as in Cones pairs. This can be attributed to the fact that same level decomposition of CT generates more subbands than those of MRSVD. Multiple subband searching, although effective for accuracy improvement, may lead to *more choice for disparity* as the search range goes higher. This apparent ambiguity results in relatively less accurate disparity estimation in higher disparity ranges for CT.

The results of segWinApprox are quite close to segWin results with improper edge restoration and reduction in minute details because approximate bands do not contain the detail parts and the edge or minute details of an image are lost. As the level of decomposition increases, the result deteriorates due to the higher resolution approximate that progressive reduce information. As the *initial disparity estimation*

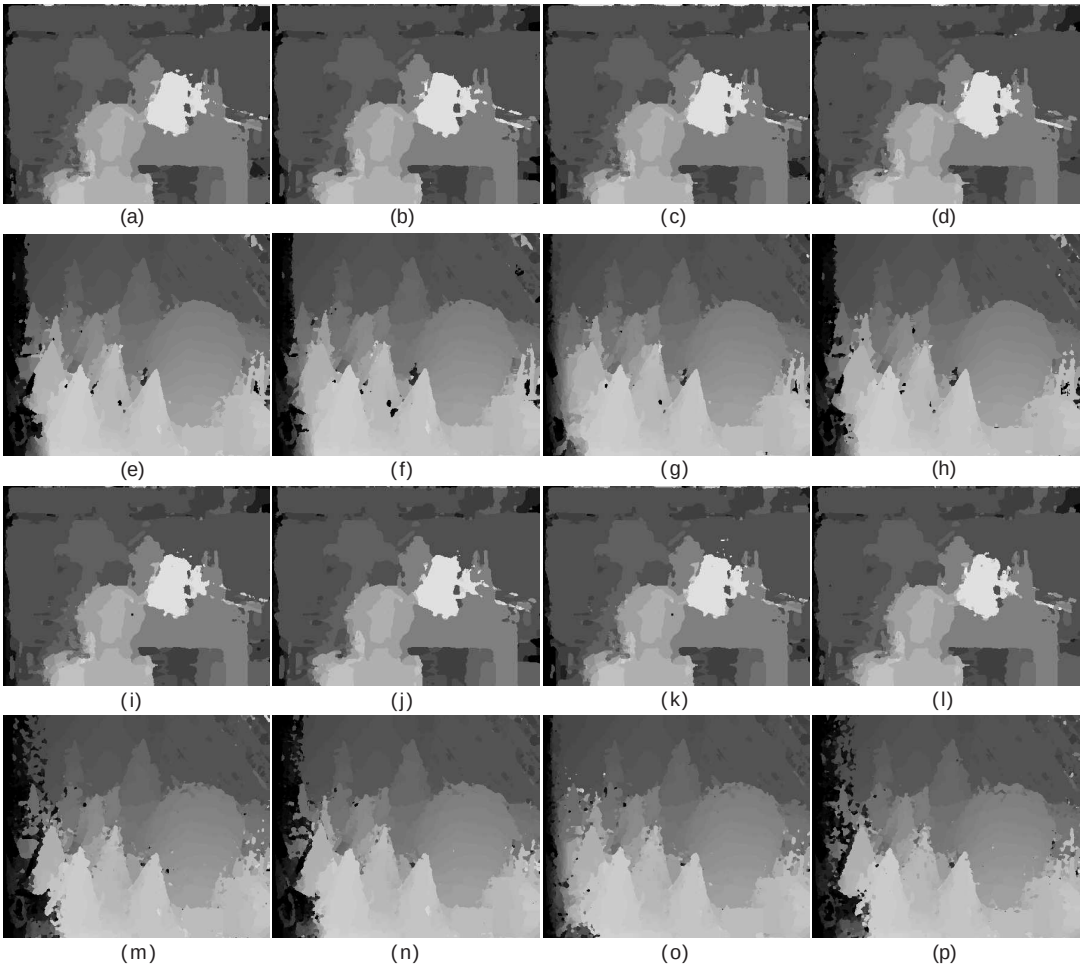

Figure 6.3: Disparity maps generated using segWinApprox. (a)-(h) represent level 1 and (i)-(p) represent level 2 decomposition. (a),(e),(i),(m) WT, (b),(f),(j),(n) CT, (c),(g),(k),(o) MRSVD and (d),(h),(l),(p) CONT.

is done in the *smallest* approximate band at highest level of decomposition, the initial map loses detailed information. The refinements in lower level approximations are unable to provide details. Thus, higher level of decomposition normally reveals less accurate disparity maps.

Chapter 7

Conclusion

The thesis is concluded with a summary of the works, the contributions and a scope for future works.

7.1 Summary of the work

The thesis begins with an introductory description of stereo vision with the related challenges involved in disparity estimation, the motivation of the work, the problem statement, the objective of the thesis and the scope of this work. The description is followed by the related literature review consisting of the general workflow of the process, the classification of stereo correspondence, the advantages and disadvantages of the different methods and the introduction of multiresolution in stereo correspondence. A brief introduction to the multiresolution analysis methods used in the work is presented next. Then, two methods have been proposed for correspondence estimation using multiresolution analysis. The first method involves curvelet decomposition with modified adaptive support weights for disparity estimation. The second method is an enhancement to the first method using curvelet decomposition, support weights and disparity calibration. The proposed methods are followed by a novel comparison framework. The comparison framework is proposed to evaluate and compare different multiresolution analysis methods for stereo correspondence.

7.2 Contributions of the Research

The contribution of the work can be divided into three distinct parts.

A novel local method for stereo-matching has been provided that uses Curvelet Transform and Modified Adaptive Support Weights for progressive improvement of disparities in multiple scales and orientations. Extensive results show that the output is better than other local methods. The method can also be used as a preprocessing step for a global method.

A highly accurate extension of the aforesaid method is provided that includes Disparity Calibration to smooth the initial disparity map using the continuity constraint. Results show that the method is comparable to the top global methods in literatures, and it can be used as a complete correspondence searching algorithm.

Finally, a comprehensive comparison framework has been proposed for evaluation of multiresolution methods in stereo-matching. Four different multiresolution methods have been evaluated and compared with extensive qualitative and quantitative results.

7.3 Future Works

The next step can be an effort to merge multiresolution to global methods to increase the performance further. Another good approach would be to include the other constraints in the proposed methods to prevent false matches.

The future work for comparison should include other multiresolution methods that are not covered in the limited scope of work. It would be a good idea to include more datasets to understand the variation in the performances of different methods and to help select methods for specific needs.

References

[1] http://vision.middlebury.edu/stereo/. [10]

[2] A. Bhatti and S. Nahavandi. Depth estimation using multiwavelet analysis based stereo vision approach. *International Journal of Wavelets Multiresolution and Information Processing*, 6(3):481–497, 2008. [21]

[3] A.F. Bobick and S.S. Intille. Large occlusion stereo. *International Journal of Computer Vision*, 33(3):181–200, 1999. [19]

[4] Y. Boykov, O. Veksler, and R. Zabih. A variable window approach to early vision. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 20(12):1283–1294, 1998. [18]

[5] P.J. Burt and E.H. Adelson. The laplacian pyramid as a compact image code. *IEEE Transactions on Communications*, 31(4):532–540, 1983. [29]

[6] E. J. Candes. What is... a curvelet? *Notices of American Mathematical Society*, 50, 2003. [26, 28]

[7] E. J. Candes, L. Demanet, D. L. Donoho, and L. Ying. Fast discrete curvelet transforms. *In Multiscale Modelling and Simulation*, 5:861–899, 2006. [27]

[8] E. J. Candes and D. L. Donoho. Curvelets - a surprisingly effective nonadaptive representation for objects with edges, 2000. [27, 28]

[9] G. Caspary and Y. Y. Zeevi. Wavelet-based multiresolution stereo vision. *In Proceedings of the 16th International Conference on Pattern Recognition*, 3:680–683, 2002. [21]

[10] Y. Deng and X.Y. Lin. A fast line segment based dense stereo algorithm using tree dynamic programming. *In European Conference on Computer Vision*, pages 201–212, 2006. [42]

[11] H. Ding, M. Fu, and M. Wang. Shift-invariant contourlet transform and its application to stereo matching. *In Proceedings of the First International Conference on Innovative Computing, Information and Control*, pages 87–90, 2006. [21, 29]

[12] M.N. Do. *Directional multiresolution image representations*. PhD dissertation, Swiss Federal Institute of Technology, Lausanne, Switzerland, December 2001. http://www.ifp.uiuc.edu/minhdo/publications. [29]

[13] A. Fusiello, V. Roberto, and E. Trucco. Efficient stereo with multiple windowing. *In Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, pages 858–863, 1997. [19]

[14] W.E.L. Grimson. Computational experiments with a feature based stereo algorithm. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 7(1):17–34, 1985. [13]

[15] Z. Gu, X. Su, Y. Liu, and Q. Zhang. Local stereo matching with adaptive support-weight, rank transform and disparity calibration. *Pattern Recognition Letters*, 29(9):1230–1235, 2008. [42, 53]

[16] H. Hirschmuller and D. Scharstein. Evaluation of cost functions for stereo matching. *In IEEE Conference on Computer Vision and Pattern Recognition*, pages 1–8, 2007. [39]

[17] A. Hosni, M. Bleyer, M. Gelautz, and C. Rhemann. Local stereo matching using geodesic support weights. *In IEEE International Conference on Image Processing*, 2009. [42, 53]

[18] R. Kakarala and P. Ogunbona. Signal analysis using a multiresolution form of the singular value decomposition. *IEEE Transactions on Image Processing*, 10(5):724–735, 2001. [28]

[19] T. Kanade and M. Okutomi. A stereo matching algorithm with an adaptive window: theory and experiments. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 16(9):920–932, 1994. [18, 19]

[20] S.B. Kang, R. Szeliski, and C. Jinxjang. Handling occlusions in dense multi-view stereo. *In Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, 1:103–110, 2001. [19]

[21] J.C. Kim, K.M. Lee, B.T. Choi, and S.U. Lee. A dense stereo matching using two-pass dynamic programming with generalized ground control points. *In Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, 2:1075–1082, 2005. [18]

[22] Y.S. Kim, J.J. Lee, and Y.H. Ha. Stereo matching algorithm-based on modified wavelet decomposition process. *Journal of Pattern Recognition*, 30(6):929–952, 1997. [21]

[23] V. Kolmogorov and R. Zabih. Computing visual correspondence with occlusions via graph cuts. *In International Conference on Computer Vision*, pages 508–515, 2001. [19, 42]

[24] C. Lei, J. Selzer, and Y.-H. Yang. Region-tree based stereo using dynamic programming optimization. *In Proceedings of the 2006 IEEE Conference on Computer Vision and Pattern Recognition*, pages 2378–2385, 2006. [19, 42, 53]

[25] S.G. Mallat. Wavelets for a vision. *Proceedings of IEEE*, 84(4):604–614, 1996. [21, 25]

[26] K. Mikolajczyk and C. Schmid. A performance evaluation of local descriptors. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 27(10):1615–1630, 2005. [18]

[27] K. Mikolajczyk, T. Tuytelaars, C. Schmid, A. Zisserman, J. Matas, F. Schaffalitzky, T. Kadir, and L. Van Gool. A comparison of affine region detectors. *International Journal of Computer Vision*, 65(12):43–72, 2005. [18]

[28] D.B. Min and K.H. Sohn. Cost aggregation and occlusion handling with wls in stereo matching. *IEEE Transactions on Image Processing*, 17(8):1431–1442, 2008. [42, 53]

[29] S.B. Pollard, J.E. Mayhew, and J.P. Frisby. A stereo correspondence algorithm using a disparity gradient limit. *Perception*, 14(4):449–470, 1985. [13]

[30] D. Scharstein and C. Pal. Learning conditional random fields for stereo. *In IEEE Conference on Computer Vision and Pattern Recognition*, pages 1–8, 2007. [39]

[31] D. Scharstein and R.S. Szeliski. A taxonomy and evaluation of dense two-frame stereo correspondence algorithms. *International Journal of Computer Vision*, 47(1-3):7–42, 2002. [12, 39, 59]

[32] D. Scharstein and R.S. Szeliski. High-accuracy stereo depth maps using structured light. *In IEEE Conference on Computer Vision and Pattern Recognition*, 1:195–202, 2003. [39, 59]

[33] C. Schmid, R. Mohr, and C. Bauckhage. Evaluation of interest point detectors. *International Journal of Computer Vision*, 37(2):151–172, 2000. [18]

[34] C. Schmid and A. Zisserman. The geometry and matching of lines and curves over multiple views. *International Journal of Computer Vision*, 40(3):199–233, 2000. [18]

[35] M. Shim. Wavelet-based stereo vision. *Proceedings of the First IEEE International Workshop on Biologically Motivated Computer Vision*, page 326335, 2000. [21]

[36] J. Sun, Y. Li, S.B. Kang, and H.Y. Shum. Symmetric stereo matching for occlusion handling. *In Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, 2:399–406, 2005. [42, 53]

[37] O. Veksler. Stereo correspondence with compact windows via minimum ratio cycle. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(12):1654–1660, 2002. [19]

[38] O. Veksler. Fast variable window for stereo correspondence using integral images. *In Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, 1:556–561, 2003. [18]

[39] L. Wang, M. Liao, M. Gong, R. Yang, and D. Nister. High-quality real-time stereo using adaptive cost aggregation and dynamic programming. *In Proceedings of the Third International Symposium on 3D Data Processing, Visualization, and Transmission*, pages 798–805, 2006. [19]

[40] Y. Wang. Principles and applications of structural image matching. *Journal of Photogrammetry and Remote Sensing*, 53(3):154–165, 1998. [18]

[41] Z.F. Wang and Z.G. Zheng. A region based stereo matching algorithm using cooperative optimization. *In IEEE Conference on Computer Vision and Pattern Recognition*, pages 1–8, 2008. [19, 42, 53]

[42] Q. Yang, L. Wang, R. Yang, H. Stewénius, and D. Nistér. Stereo matching with color-weighted correlation, hierarchical belief propagation, and occlusion handling. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 31(3):492–504, 2009. [19, 42, 53]

[43] Q.X. Yang, L. Wang, and R.G. Yang. Real-time global stereo matching using hierarchical belief propagation. *In British Machine Vision Conference*, pages 989–998, 2006. [19, 42]

[44] K. Yoon and I. Kweon. Locally adaptive support-weight approach for visual correspondence search. *In IEEE Conference on Computer Vision and Pattern Recognition*, 2:924–931, 2005. [34, 38, 42, 53]

[45] W. Zhang, Q. Zhang, L. Qu, and S. Wei. A stereo matching algorithm based on multiresolution and epipolar constraint. *Proceedings of the Third International Conference on Image and Graphics*, page 180183, 2004. [21]

Vita Auctoris

NAME	:	Dibyendu Mukherjee
BIRTH YEAR	:	1983
BIRTH PLACE	:	INDIA
EDUCATION		
2008–2010	:	**Masters of Applied Science**
		Electrical and Computer Engineering
		University of Windsor, Windsor, Ontario, Canada
2002–2006	:	**Bachelors of Engineering**
		Electronics and Telecommunications Engineering
		Bengal Engineering and Science University, Howrah, India