Towards a Robust, Passive Stereo Depth Sensor with Confidence and Intensity Guided Anisotropic Diffusion Disparity Refinement

Matthew Taylor Dooner
Bachelors of Science, Computer Science

Submitted in fulfilment of the requirements for the
Degree of Masters by Research in Computing Science

Department of Computing Science
School of Science and Engineerings
University of Glasgow

September 2012

Abstract

Stereo matching is the most common method for recovering depth information from two dimensional images. Despite the large amount of attention given to the problem it remains unsolved, and even robust methods sometimes produce noisy and inaccurate depth measurements. Existing disparity refinement methods can filter this output noise as a post-processing step at the cost of some fine depth detail. This work establishes a method to reduce noise while preserving the two-dimensional structure of the image through a modification of the well-known anisotropic diffusion technique. Weighting the amount of diffusion based on the edge strength of the intensity image rather than the edge strength of the disparity preserves a greater number of depth boundaries. The confidence of the disparity estimate prevents diffusing bad estimates into good estimates and creates a stopping criteria for the diffusion process. Varied datasets provide validation of the technique; a dataset of our own design combined with two established benchmark datasets test the algorithm in varied environments. The performance of the author's technique is compared against the technique which it improves on and the most closely related technique from recent literature. The author's Confidence and Intensity Guided Anisotropic Diffusion (CIGAD) outperforms the other techniques in many cases and provides more reliable and robust results overall.

Contents

List of Tables

List of Figures

List of Algorithms

1 Introduction

1.1 Problem Description

This Masters by Research advances the state of computational stereo matching with an extension of the well-known anisotropic diffusion image regularisation technique. The stereo vision process is superficially similar to how humans and other mammals perceive depth from two eyes; the spatial offset in the projection of an object onto two imaging planes reveals the depth of the object. In a computer stereo vision system the projection from one imaging plane to the other is determined computationally. The stereo matching algorithm computes a disparity map, the projection of each point in one image to its corresponding location in the second image. The two input images are information-rich with a relatively high signal-to-noise ratio, but the output disparity has a low signal-to-noise ratio caused by ambiguities inherent in the matching process. Many techniques use smoothing in a regularisation step to reduce this noise, but smoothing results in a loss of sharp features and can preserve or propagate large errors to decrease the robustness of the result. The author develops a technique that adapts anisotropic diffusion to take into account the intensity image structure and matching confidence to avoid these detrimental artefacts of smoothing. By exploiting tendency for intensity image edges to predict the occurrence of depth edges and by discounting the less-confident disparity estimates the Confidence and Intensity Guided Anisotropic Diffusion (CIGAD) technique produces a more robust disparity map.

The CIGAD process was chosen based on two hypotheses: inhibiting diffusion over intensity edges maintains the depth map structure and filling low-confidence regions with disparity estimates from neighbouring high-confidence regions reduces noise and suppresses erroneous results. Maabar [30] develops the intensity-guided technique as a multi-resolution post processing step. Since most computational stereo algorithms already include a regularisation step to enforce the continuity constraint, the author chooses to embed the technique within the stereo matching algorithm to preserve more of the depth map structure and avoid computing the regularisation twice. The removal of the extra regularisation step and the possibility of a parallel implementation make the author's technique more suitable for robotics applications.

1.2 Objectives and Motivation

The goal of this MRes is to develop a more robust stereo matching algorithm with an application in robotics. The algorithm seeks to improve the robustness and quality of the stereo matching results by suppressing spurious data and reducing the noise level of outputs. The EU framework project CloPeMa will use the new algorithm as part of the robot perception capabilities for a cloth manipulating robot. The motivation for this project came from a need for more robust, passive remote sensing discovered during the authors' previous work in robotics and the author's involvement with the

CloPeMa application process.

For robots to be generally useful they must operate in generalised environments. I learned this firsthand while competing in the 2005 and 2007 DARPA Challenge autonomous vehicle races [46, 45] which accelerated research in driverless cars. While autonomous robots had been operating in controlled settings for decades and field robotics was able to remotely control rovers on Mars, researchers had not been widely pursuing fully autonomous robots for tasks of this scale. I believe the important advances in robotics are made when the inherent uncertainty of real-world situations must be dealt with. Similarly, the CloPeMa project aims to enable robotic manipulation of cloth in less structured environments.

The acronym CloPeMa stands for Cloth Perception and Manipulation and is part of a European Commission Seventh Framework Programme for small projects. This project plans to "to advance the state of the art in the autonomous perception and manipulation of all kinds of fabrics, textiles and garments". Previous attempts at cloth manipulation have required specially constructed working environments for the robot and could only manipulate limited types of garments. CloPeMa will sort and fold a pile of varied garments in a more arbitrary setting. The five partners are the Centre for Research and Technology Hellas (CERTH), University of Glasgow (UoG), Czech Technical University in Prague (CVUT), Universita degli Studi di Genova (UniGe), and Neovision s.r.o. An industrial two-arm robot manipulator with standard grippers will be purchased, UniGe will build a more sophisticated tactile gripper, and the University of Glasgow will build an active stereo head. CERTH will focus on planning and photometric stereo, and CVUT's expertise is in compliant control. My work on disparity map regularisation will increase the robustness of UoG's active stereo head work package.

1.3 Contributions

During his studies and research the author contributed;

- An experiment demonstrating why the Maabar and Siebert disparity map regularisation technique cannot be directly applied to the C3D stereo matching algorithm.
- An extension of Maabar and Siebert's technique that can be applied to the C3D stereo matching algorithm to produce robust and noise reduced disparity maps.
- A case study validating the applied techniques.

1.4 Dissertation Structure

This dissertation first reviews the important concepts and recent work in stereo matching and disparity refinement in Chapter 2. Next the topic of anisotropic diffusion [33] is

similarly summarised in Chapter 3. These two topics, stereo matching and anisotropic diffusion, form the foundations of the author's work. Chapter 4 outlines the two disparity map refinement techniques and describes how they embed into the stereo matching process. Finally, Chapter 5 evaluates the developed disparity map regularisation techniques in the context of the author's thesis.

1.5 Thesis Statement

A hierarchical stereo matching technique with a novel disparity refinement step robustly recovers depth information from a stereo pair of images for the robotic cloth manipulation application. Disparity refinement based on both confidence and intensity image information improves the reliability of the stereo matching algorithm. Retaining information about the two-dimensional structure of the scene prevents the blurring of depth boundaries, and the confidence information promotes filling of regions with matcher errors from neighbouring regions with correct disparities.

2 Literature Review of Stereo Correspondence and Disparity Refinement

2.1 Introduction

The goal of stereo matching is to find where the corresponding the locations of objects captured in one view project into another. The process begins when two cameras capture images of the same scene from two different viewpoints. This configuration is described by the stereo camera and epipolar geometry, and the geometric relationships constrain the possibilities for the projections between the two views. The stereo matching algorithm computes a mapping between the two views called a disparity map. The entire matching process can be divided into four stages: matching cost computation, cost aggregation, disparity computation, and disparity refinement. Occlusions, noise, and match failure all contribute to mistakes in the first three stages, and the last stage, refinement or regularisation, attempts to remedy these errors. Finally, some stereo matching techniques compute disparities at different scales to take advantage of the scale space concept from the signal processing field.

2.2 Image Capture Hardware

The stereo capture rig (Figure 1) adopted in this work uses two Prosilica GC2450 ethernet cameras (one colour and one monochrome) mounted on a robotic binocular stereo head. The cameras are mounted on motor-actuated turn-tables to control their pitch and yaw. Each camera is manually centred and focused on the subject and not moved during the capture process. Non-flash fluorescent umbrella soft-boxes provide adequate lighting. A PC triggers each image capture and immediately downloads the images from the cameras. This robotic stereo head captures the mannequin head dataset used to evaluate the improved disparity map regularisation process. A blue backdrop helps segment the mannequin head from the background.

2.3 Stereo Camera Geometry

Stereo matching uses two views of a scene to calculate the depth of objects in the scene. The views and objects are related by the camera and epipolar geometry (Figure 2). The object P is projected onto each camera's imaging plane (L or R) through the focal origins O_L and O_P. The ray between P and O intersects the imaging plane at point p, and the amount of light reflected from P determines the pixel intensity at p. The depth of an object at P is the length of $\overline{OP}$ and the focal length of the camera is the distance between the focal plane and O. The epipolar line $\overline{pe}$ determines the possible locations in the image plan of a point at varying depths.

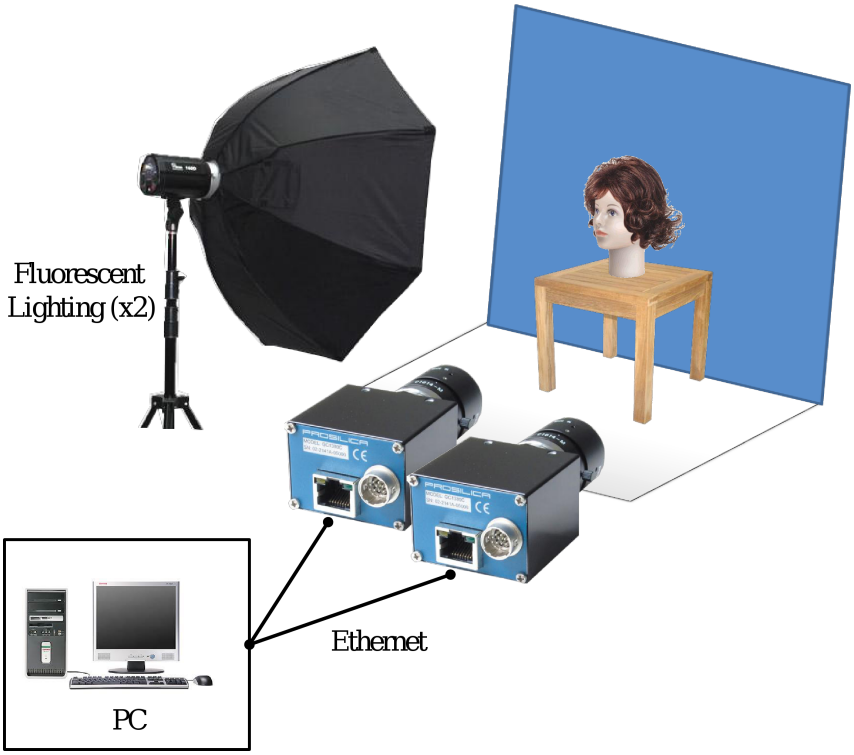


Figure 1: A diagram of the Computer Vision and Graphics Group's robotic stereo head capture system.

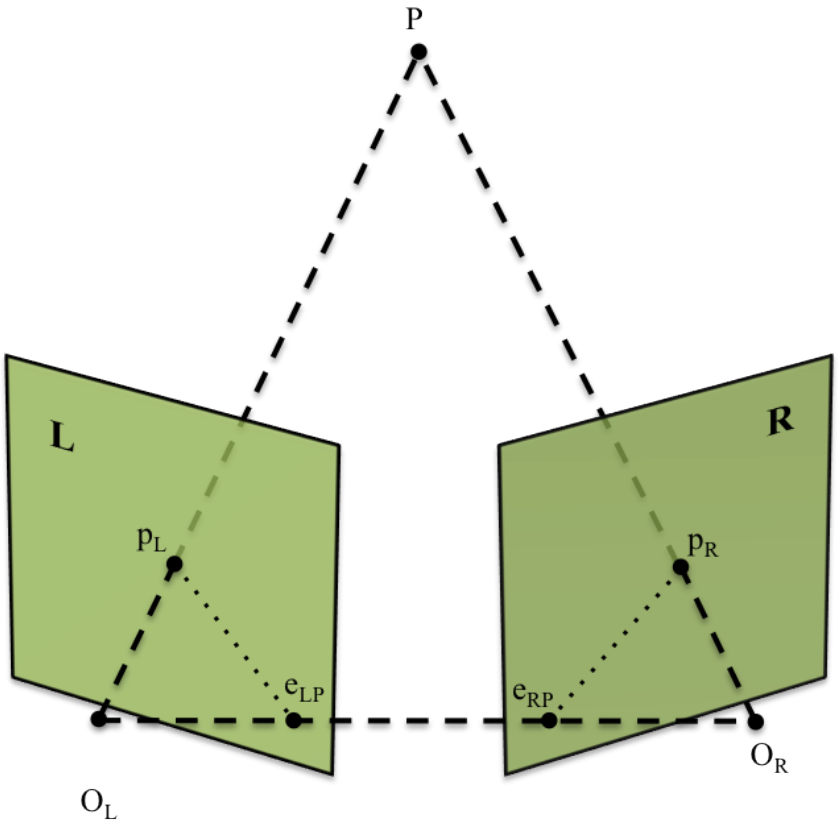


Figure 2: The epipolar geometry for two views L and R with projective centres O_L and O_R determines where object P is projected onto the image places while varying depths P_0 and P_1.

2.4 Disparity

The problem of stereo correspondence is defined for two images captured from slightly different viewpoints of the same scene; determine the mapping from each pixel p_L in one image to its corresponding location p_R in the other image. These mappings are called disparity and are typically represented in a matrix of vector tuples (d_x, d_y). The disparity maps generally have the same number of rows and columns as the image. The disparities can be whole numbers of pixels but are commonly interpolated to fractions of pixels with sub-pixel refinement. A scalar disparity map (either d_x or d_y is fixed at 0) represents a fronto-parallel scan-line-aligned camera configuration where the epipolar lines are horizontal and parallel. The camera calibration can be used to re-project the input images to mimic this setup. When the images are scan-line-aligned the disparity search only occurs along the rows of the image because the mappings occur along the horizontal and parallel epipolar lines. Disparities determined from searches along the images rows are called horizontal disparity. When the input images are not scan-line-aligned the disparities are vector-valued in two dimensions because the disparity search must also occur along the image rows. This second dimension is called vertical disparity.

2.5 Occlusions

Not all pixels have valid disparity values; when part of the scene is visible in one view but not the other no such mapping exists. These regions are called occlusions, and there are several causes for them. A portion of the image border on one or two adjacent sides will always be occluded due to the geometry and offset of the stereo imaging apparatus. Next, an object can occlude itself (self-occlusion) when part of that object is only visible from one of the viewpoints. These facets that are only visible from one view will have undefined disparities. In Figure 18 the mannequins left cheek self-occludes her left ear. Finally, a foreground object can occlude a background object differently in two views due to the camera perspective. Figure 3 demonstrates this because the block in the background is not occluded in the left but but is partially occluded by the cube in the right view. The right edge of the block has an undefined disparity. In a disparity map, occlusions are often denoted by special values or an occlusion mask.

2.6 Noise and Match Failure

Two types of error occur in disparity maps: noise and match failure. Noise is the random and meaningless fluctuation in the disparity values caused by fluctuations in the underlying intensity images. This type of error is caused by the imaging sensor, and its primary causes are quantisation and Shott noise (photon counting) [15]. Two types of quantisation, spatial and colour, contribute to quantisation noise. Spatial quantisation occurs as the result of the discretisation of the image plane into pixels.

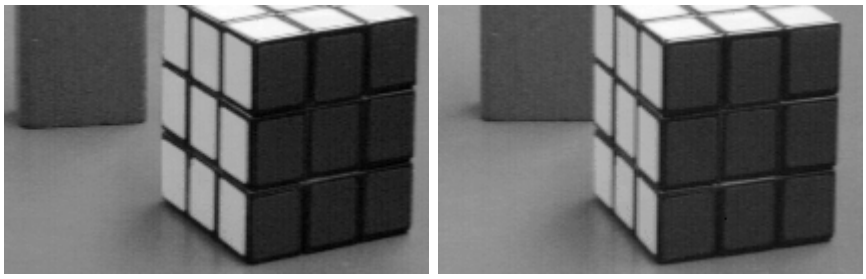


Figure 3: A Rubiks cube in the foreground occluding a block in the background. This image pair is the "toy" dataset from the USC Institute for Robotics and Intelligent Systems Stereo Image Data Base [14].

The edge of an object that bisects the area of a single pixel contributes to the light information captured over the entire area of the pixel. Colour quantisation refers to how this light information is stored. The brightness of a pixel is represented with a fixed number of bits. 8, 16, and 24 bit images are most common. Negligible colour discretisation occurs in a 24-bit monochrome sensor, but quantisation effects are significant with colour sensors at this depth [15]. Colour sensors usually use a Bayer pattern colour filter array so that each sensor cell measures either red, green, or blue light. To construct a complete image the colour of each pixel is interpolated from a region including its neighbours, reducing precision with demosaicing effects. The two main types of imaging sensors, CMOS and CCD, are used in this dissertation. Historically crosstalk between neighbouring photon collectors was an additional source of sensor noise in CCD sensors. Modern imaging sensors have nearly eliminated this problem and there is little difference in quality between CMOS and CCD sensors [15].

Match failure is the systemic choice of incorrect disparity hypotheses within a region of the image. Two possible input image causes of match failure include wide texture-less areas and occluded regions. Texture-less regions contain little information that differentiates one pixel from its neighbour and can skew or induce noise in the disparity measurements. A texture-less region would provides no means to discern between the two disparity estimates P_0 and P_1 in Figure 2. Large amounts of sensor noise can produce the same effect as texture-less regions when the noise level grows larger than the signal strength. Occlusion can also introduce ambiguity to the matching process. Determining which regions of the scene are occluded is a non-trivial problem and algorithms usually attempt to estimate disparity in these regions even though the actual value is undefined. Any match in an occluded region will always be incorrect and contribute to disparity map errors, but it is often desirable to fill the occluded region with an appropriate disparity.

2.7 Regularisation

Regularisation applies to a wide range of mathematical problems and generally refers to an additional term that prevents over-fitting a model to underlying data. In the case of stereo matching regularisation prevents fitting disparities to noise in the input pair of

images. In this application regularisation is also referred to as the continuity constraint or smoothing. The main body of this dissertation is concerned with reducing the effects of errors from noise and occlusions using adaptations of the anisotropic diffusion image regularisation technique. Disparity map regularisation is discussed in detail in Section 2.

2.8 Stereo Matching

Computational stereo correspondence is a mature but still developing area of stereo vision that began in the early 1970's, and a review of this early work was conducted by Barnard and Fischler [6]. Brown et. al. summarise more contemporary techniques with an emphasis on methods that handle occlusions [10]. A more contemporary review by Scharstein and Szeliski [40] divide the stereo correspondence problem into four steps: matching cost computation, cost aggregation, disparity computation, and disparity refinement. Modern textbooks on the subject [15] also use this convention. The state of the art continues to be advanced using the Scharstein and Szeliski datasets and evaluation method. Image pyramids, a scaffolding upon which disparity computation can be performed, is also discussed.

The applications for stereo correspondence are wide and varying. The field has its roots in photogrammetry whose techniques date back to the early days of photography. Computational stereo is commonly used today for land surveys from aircraft or satellite. The technique also has a number of applications in robotics. For industrial robots it can facilitate part picking and alignment. Stereo is used in mobile robotics for navigation and obstacle avoidance. Manufacturing can perform automated quality assurance on parts by measuring them with stereo vision systems. By using more than one view, entire objects, buildings, or cities can be mapped three dimensionally. Medicine uses stereo to measure topographical changes in patients and to track and measure tissues in robotic surgery. Stereo is widely considered to be a special case of optical flow which can support object and person tracking, change detection, video compression, and ego motion.

The body of work covering computational stereo is impractical to review in its entirety, so only the portions related to the work of this dissertation will be covered in this literature review. This report will cover multi-scale or hierarchical approaches, algorithms that include a diffusion component, and finally the C3D algorithm used in the experiments.

2.9 Matching Cost Computation and Aggregation

The matching cost function establishes hypotheses for the disparity of each pixel in a stereo pair of images. In practice, the cost function estimates how similar the neighbourhood containing a pixel in one image is to a neighbourhood in another. Many different cost functions have been proposed [15]. Two of the most popular, the sum of

absolute differences (Equation 1) and the sum of squared differences (Equation 2), are often used for their computational efficiency [10]. More recently mutual information (MI) has gained popularity as a matching cost [27]. Both SAD and SSD sum the pixel intensity difference over a window of size $[i, j]$ centred at pixel (x, y).

$$C_{SAD}(x,y) = \sum_{i,j} |I_R(x+i, y+j) - I_L(x+i+d_h(x,y), y+j+d_v(i,j))| \quad (1)$$

$$C_{SSD}(x,y) = \sum_{i,j} \left[I_R\left(x+i, y+j\right) - I_L\left(x+i+d_h(x,y), y+j+d_v(x,y)\right)\right]^2 \quad (2)$$

The cost aggregation step simply combines the costs of the neighbours to produce a more robust estimate. Aggregation is often part of the cost function as seen in the summation over the ranges of i and j in Equations 1 and 2. The range (window size) is usually fixed, but this can cause depth edges to be blurred or corners to be rounded off when these edges occur at scales the window size does not respond well to. Windows with an adaptable size have been proposed [26] but this problem is also solved using hierarchical matching as described in Sections 2.13 and 2.14. Matching algorithms that use a global optimisation step to find the disparities often do not have an explicit aggregation step, but the smoothness term imposed by many of these methods perform the aggregation role.

2.10 Disparity Computation and Sub-pixel Disparities

Search window or block-based matching such as those discussed in Section 2.9 usually employs a winner takes all (WTA) approach where the disparity that has the lowest matching cost is chosen. Methods that attempt to choose a disparity that also minimise the cost at neighbouring pixels are called *global optimisation* methods. Some examples include dynamic programming [22], graph cuts [9], energy minimisation [39], and local support [19, 20]. Combining multiple approaches can yield more robust matching results [16, 17, 29].

Most disparity assignment techniques compute whole pixel disparities which translate into blocky depth maps and suffer from aliasing errors. Many algorithms compute subpixel disparities by finding the maximum of a quadratic curve fitted to integer disparities. Another option is to perform the subpixel interpolation as part of the matching process. For example instead of sweeping x and y from Equation 1 over whole numbers to compute the matching cost fractional steps could be taken instead. The pixel values must be proportionally interpolated from their neighbours, adding to the computational complexity of the algorithm. Some summaries treat subpixel interpolation as a refinement process [40], but this dissertation (Section 2.12) defines disparity refinement as method methods that reduce noise or correct errors in disparity maps. Matching algorithms often estimate the errors in disparity computation using a metric called *confidence*.

2.11 Confidence

The confidence metric estimates the errors in the computed disparities. Hu and Mordohai [23] provide the most recent and relevant review of confidence metrics. The authors focus on local confidence metrics which are inversely proportional to the disparity error and that can predict occluded regions. The C3D matching algorithm [25] (Section 2.15) used in this dissertation for baseline measurements produces a confidence metric that is consistent with these assumptions. Hu and Mordohai claim that three factors can contribute to a confidence metric; the matching cost, the matching cost curve/surface, and the left-right consistency. The matching cost indicates the strength of the match because a hypothesis with a very low cost suggests a very similar neighbourhood of pixels in the left and frames at the measurement site. Hu and Mordohai use a method called the Matching Score Metric (MSM) which assigns the confidence to the negative minimum matching cost c_1:

$$C_{MSM} = -c_1 \tag{3}$$

C3D uses the MSM method to compute the confidence for the baseline matches in this dissertation. One problem with MSM is that it can falsely report a strong confidence estimate for textureless regions. Textureless regions produce a high confidence despite a large match uncertainty because the confidence depends only on the intensity values in the neighbourhood of the match. To avoid this problem the confidence metric uses the first (c_1) and second (c_{2m}) local cost minima. A large ratio indicates a distinct match while a small ratio indicates two or more similar hypotheses. Hu and Mordohai summarise several versions of this approach into a metric named Peak Ratio (PKR):

$$C_{PKR} = \frac{c_{2m}}{c_1} \tag{4}$$

The PKR method performs the best overall in Hu and Mordohai's evaluation. Another popular disparity refinement, the Left Right Consistency (LRC), exploits the possibility of matching the input frames by both left against right and right against left. Although the disparity from the right should yield an inverse mapping of the disparity from the left, the same factors that contribute to error and noise in the disparity estimates contribute to discrepancies in these calculations. LRC is the difference between a disparity measurement in the left image $(d_x, d_y) = D_L(x, y)$ and the disparity of the pixel in the right image mapped to by D_L:

$$C_{LRC} = \Big|(d_x, d_y) - D_R(x + d_x, y + d_y)\Big| \tag{5}$$

LRC performs better than MSM, and it is possible to compute LRC with little extra computational complexity. By storing the costs when calculating the disparities in one direction, it is possible to then traverse the costs in the reverse direction without recomputing the match hypotheses. Many algorithms apply a threshold to LRC and reject disparity estimates with a confidence greater than the specified range. The

Banno and Ikeuchi disparity refinement technique [5] uses LRC for confidence and is evaluated in this dissertation.

2.12 Disparity Refinement

The disparity refinement step follows disparity computation and attempts to reduce noise or otherwise improve on the initial estimates. The disparity refinement step fills up to three roles: regularisation, occlusion filling, and interpolation (although occlusion filling can be considered a special case of interpolation). The regularisation role filters out small, spurious variation to reduce the overall noise of the disparity map. Occlusion filling provides disparity estimates for these areas in which the disparity is undefined. Typically the occluded region is filled with disparities from the background. Finally, interpolation is important for matching processes that do not produce dense matches. If the matcher rejects disparities with a low confidence, the interpolation role estimates an approximation to the correct disparities from the local neighbourhood.

The disparity refinement step often incorporates local information from the local neighbourhood near each measurement and a confidence metric. Two classical and popular local disparity refinement techniques are the median filter and the Gaussian convolution. The Gaussian convolution combines a disparity estimate with those of its neighbours according to weights defined by a Gaussian distribution. This method mainly reduces the noise in the disparity map, but the Gaussian filter also reduces the amount of fine detail available in the final disparity map. Diffusion performs a function similar to a Gaussian convolution, but an adaptation called anisotropic diffusion solves problems of the Gaussian convolution destroying edges and fine detail. For a complete discussion of diffusion's roles in disparity refinement refer to Section 3.

Models or surfaces can be fitted to the disparity measurements [44] to reduce noise and uncertainty by imposing structure to the scene. One problem with fitting fixed shapes such as planes is that the shapes do not fit the surface, creating similar artefacts as introduced by isotropic blurring. Banno and Ikeuchi note poor performance on real-world data [5] with plane fitting approaches that perform well on piecewise-planar scenes such as the Middlebury dataset (as reported in Szeliski's book chapter [43]). Polynomial patch and spline registration can better match the surface, but an over-fitted surface introduces more error to the disparity map. Patches and splines also cannot inherently handle depth discontinuities.

2.13 Scale Space

Hierarchical techniques decompose an image into a number of frequency, or scale levels. This differs from the Fourier transform which decomposes an image into the frequency domain because the outputs remain in the spatial domain required for local image matching. Witkin proposed the first scale-space approach based on filtering [48]. A Gaussian filter is repeatedly applied to a signal. With each application the highest

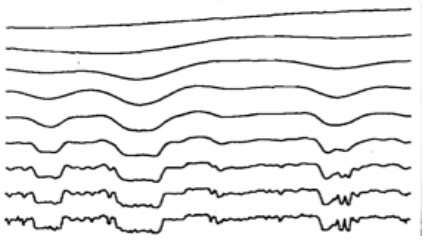


Figure 4: One-dimensional view of the two-dimensional scale space proposed by Witkin [48] incrementally filters the highest frequencies in each level, starting from the bottom. Figure appears in [48].

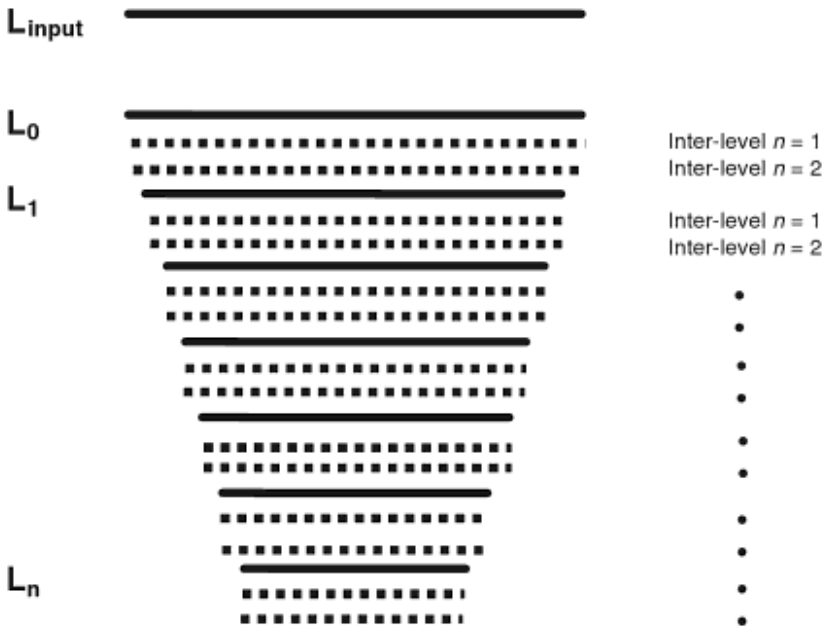


Figure 5: Construction of a semi-pyramid. Illustration appears in p.180 of [15]

frequencies are removed while the lower frequencies are left in residual. A valid scale space does not create new features or add shift, or bias, to the signal [28], and spatial features trace easily through the different scales in which they occur. Babaud et al. prove that the Gaussian kernel is unique in its applicability to scale-space filtering [4] although a scale-space constructed using an alternate technique is discussed in the next section [33]. The construction of a scale-space solves the false target problem by traversing structures of different scales.

Scale-space theory becomes particularly useful when applied to an image pyramid. Burt and Adelson describe a method for constructing the Gaussian pyramid and the Laplacian pyramid [11]. The process involves repeatedly filtering and subsampling an image to create a series of images that decrease in the content of their spatial frequencies and in their physical size (Figure 6). At every level of a Gaussian pyramid all of the larger frequencies are still represented, but the Laplacian pyramid is a bandpass representation where only the current level's frequencies are represented. The Laplacian pyramid can be constructed quickly using the Difference of Gaussian (DoG) estimation [15]. To estimate the Laplacian pyramid using the Gaussian pyramid in Figure 6, each

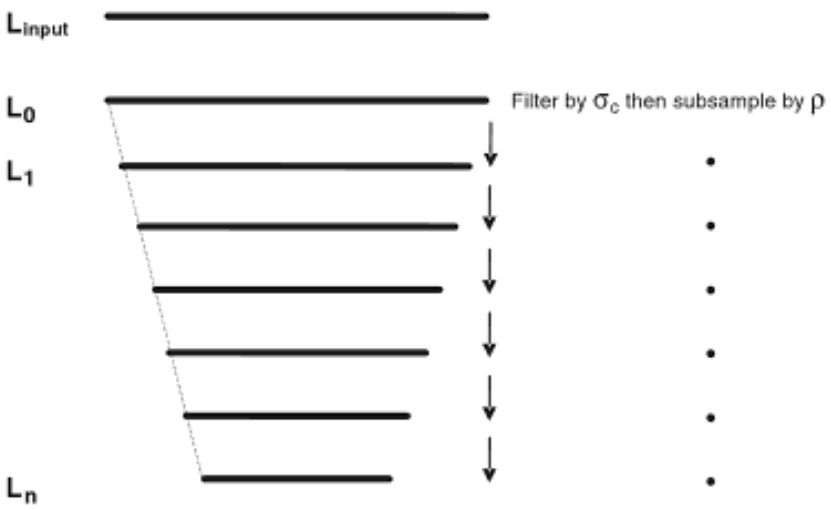


Figure 6: Construction of a Gaussian pyramid by filtering and subsampling. Illustration appears on p.175 of [15]

of the L_n levels is expanded to the size of the image at L_{n-1} and subtracted from it. This difference is a DoG image.

The benefits of constructing an image pyramid are three-fold. First is the aforementioned issue of traversing structures at different scales to apply operations in a scale-invariant manner. Secondly, the data structure is more compact than storing information about each scale at a constant resolution. Finally, using a pyramid can reduce the computational time-complexity of image processing.

2.14 Hierarchical Stereo

Modern stereo correspondence algorithms commonly use image pyramids in hierarchical methods to improve their robustness and efficiency. Correspondence algorithms are easily fooled into estimating incorrect disparities when presented with fine, repeating textures. Noise also disrupts the similarity measurement between similar image sites. For these reasons there are many sites in both images that are equally similar and thus have the same matching cost. Matching within a pyramid resolves these ambiguities at coarse scales. Next, the calculation always involves some search process to find the optimal disparity within its range. If the range of disparities is large then the search space is also large. A well constructed scale space is able to resolve the large disparities at coarse scales to limit the search range at all subsequent fine scales.

Quam introduced an early multi-scale warping stereo algorithm [35] that iteratively matches from coarse-to-fine, adding refined disparity estimates every step of the way. The disparity map is initialised to zero and process begins at the coarsest scale. The disparity map is used to warp one of the stereo pair images into correspondence with the other. Matching is performed on the warped pair and the disparities are added to the existing map. Then the disparity map is stretched to the size of the finer image in the pyramid and the process is repeated. Many algorithms use this process [44, 7] but

some do not. Fua rejects the notion that coarse scale disparities are more reliable than the fine scale ones [19]. The stereo pair is not warped into correspondence, and instead of adding the disparities at each scale the finer scale disparities replace the coarser scale ones. This works because the approach does not output disparity values when they might be ambiguous so the coarse scale disparities sometimes fill in holes left by matching in the finer scales. Hirschmuller uses the coarse scale disparities to calculate probability priors for calculating mutual information [22]. Yang inverts the process, traversing the pyramid from fine to coarse to aggregate local support for matching [49].

2.15 Multiple Scale Signal Matching (C3D)

The University of Glasgow's C3D [25] software performs the stereo matching for the experiments in this dissertation. It is a hierarchical warping correspondence algorithm that operates in scale-space on a difference of Gaussians (DoG) image semi-pyramid. A semi-pyramid (Figure 5) fills the gaps between two pyramid scales by iteratively resizing the coarser image to create a number of intermediate steps [15]. Bilinear interpolation is used to scale the intermediate images. From coarse to fine, the disparities from the coarse layer are scaled to the finer layer and the stereo pair images are warped into correspondence. The disparities are calculated for these warped images and added to the previous disparity estimates. The matching cost (Equation 6) resembles the maximum Pearson product-moment correlation coefficient. The variance and covariance, *cov* and *var* respectively, are calculated over a Gaussian window where l and r represent the left and right images and the coordinate (x, y) is the centre of the window.

$$\mathbf{cor_{l,r}}(x, y) = \frac{cov_{l,r}(x, y)}{var_l(x, y) var_r(x, y)} \tag{6}$$

Sub-pixel refinement is accomplished by both calculating costs at fractional pixel steps and by maximising a curve fitted to all of the cost estimates. After each DOG image is matched a disparity refinement step is performed to enforce the continuity constraint. Finally, the disparity and confidence maps are scaled up to the next finer resolution and the pipeline (Figure 7) continues. The process terminates after the disparity refinement step at the finest resolution of the image pyramid.

This algorithm does not require rectified input images; it determines both horizontal and vertical disparities independently of one another. For simplicity, both horizontal and vertical disparities are represented as D and each process (scale, warp, match, regularise) is performed separately in each direction. It is the last process, regularise, that the remainder of this dissertation is concerned with.

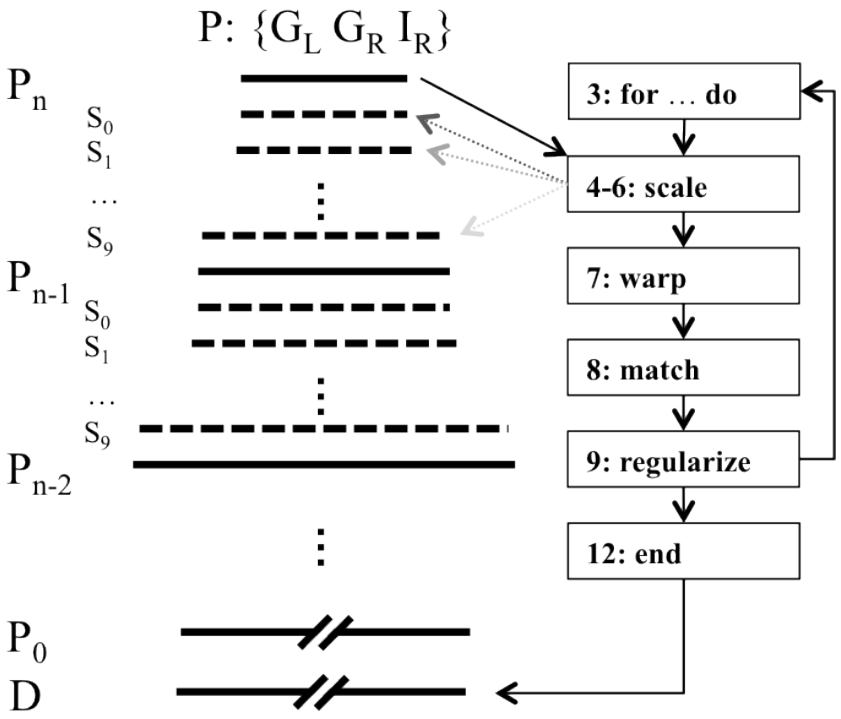


Figure 7: Each half-octave image pyramid level P_i is linearly subdivided in 10 iter-levels S_i by stretching P_i. Lines 4 through 10 of Algorithm 1 are repeated for each iter-level $S + i$ in the pyramid. D begins blank and is stretched at each iteration until the full-scale disparity map is formed.

Algorithm 1 C3D

G: Difference of Gaussians Input Image Pyramid
I: Gaussian Input Image Pyramid
D: Disparity Map
C: Confidence Map
L: Left View
R: Right View

1: $D \leftarrow D_{init}$
2: **for** $i \leftarrow \|G_L\| : -1 : 1$ **do**
3: **for** $j \leftarrow 1, numSemiLevels$ **do**
4: $L \leftarrow \textbf{scale}(G_L[i], i, j)$
5: $R \leftarrow \textbf{scale}(G_R[i], i, j)$
6: $D \leftarrow \textbf{scale}(D, i, j)$
7: $L' \leftarrow \textbf{warp}(L, D)$
8: $\begin{bmatrix} D_{tmp} & C \end{bmatrix} \leftarrow \textbf{match}(L', R)$
9: $D \leftarrow D + D_{tmp}$
10: $D \leftarrow \textbf{regularise}(R, D, C, I_R)$
11: **end for**
12: **end for**

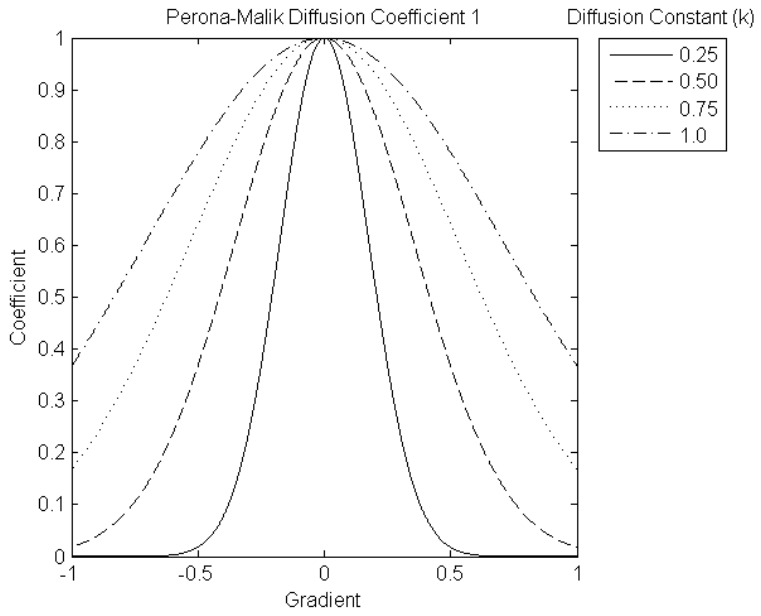


Figure 8: Plot of the Perona-Malik Diffusion Coefficient 1 (Equation 7).

3 Literature Review of Anisotropic Diffusion

3.1 Introduction

Anisotropic Diffusion is a popular image regularisation process that is able to suppress noise without blurring edges. This method is used in the author's research to increase the accuracy and robustness of computed disparity.

Perona and Malik created a paradigm shift when they introduced their seminal work on anisotropic diffusion in 1990 [33]. Not only does the technique enable feature-preserving image regularisation, it is also capable of generating scale spaces and extracting edges. Scale spaces produced by anisotropic diffusion reduce the need for scale space tracing. In a traditional scale space the location of edges and other features are shifted at coarser scales. Perona and Malik refer to Clark [13] for analysis of the two ways these "phantom edges" occur in smoothed images and for the method of tracing edges through scale space. An edge can disappear entirely at coarser scales or two edges can merge to produce a new edge at a different location from either of its parents. To correct this such features must be tracked through the hierarchy, increasing the expense of the operation. An anisotropic diffusion reduces the need for scale space tracing by preserving "the natural boundaries of objects" [33] manifested in sharp intensity gradients. While the the body of this research does not explicitly address edge detection, these concepts are important for preserving features in image regularisation processes.

Most smoothing processes to date are based on isotropic diffusion that blends pixel values in the same way in every direction or on convolution methods using kernels that responded to certain features. In contrast, anisotropic diffusion selectively diffuses along instead of across edge boundaries, preventing the common problem of blurring

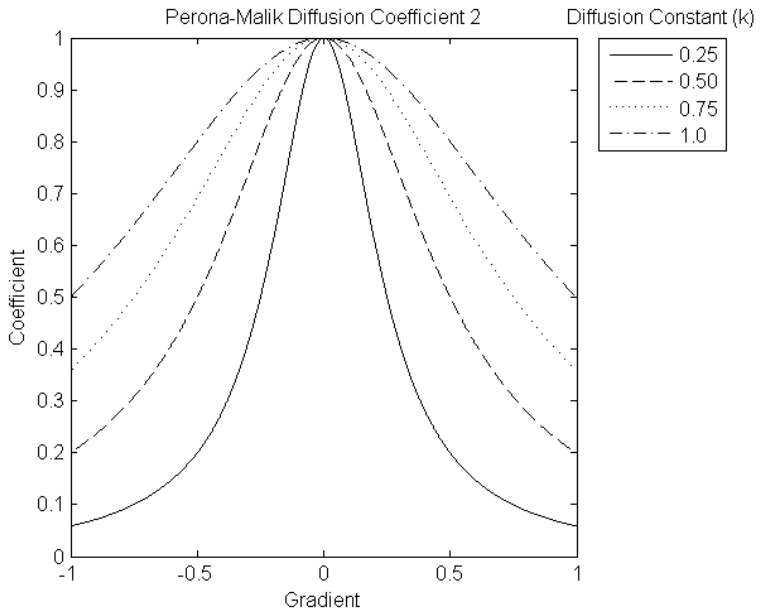


Figure 9: Plot of the Perona-Malik Diffusion Coefficient 2 (Equation 8).

of sharp features. A scale space can be constructed with this method by varying the scale constant (or diffusion constant) K, inherently segmenting the image into regions and exposing the images' edges. For image regularisation a value for K must be found such that noise is reduced while the underlying image structure is preserved. This constant controls how strong an edge must be to inhibit diffusion. The diffusion coefficient function $\mathbf{g}(\nabla I)$ maps edge strength (image gradient) to an analog of thermal conductivity [33]. Two options for the diffusion coefficient were originally proposed. The first (Equation 7) "privileges high-contrast edges over low contrast ones" and the second (Equation 8) "privileges wide regions over smaller ones" [33].

$$\mathbf{g_{pm1}}(\nabla I) = e^{-(\frac{||\nabla I||}{K})^2} \tag{7}$$

$$\mathbf{g_{pm2}}(\nabla I) = \frac{1}{1 + (\frac{||\nabla I||}{K})^2} \tag{8}$$

Diffusion is applied iteratively over a number of time steps in Equation 9. At each step finite differences in four or eight directions estimate the image (or disparity) gradients ∇I. Equation 9 only considers gradients along the four cardinal directions and ignores the diagonals. Empirical testing produced results with similar quality but greater execution times when eight directions were used instead of four. The four-direction formula is used for the remainder of this work, and gradients calculated in this manner are designated by the ∇ operator.

In each direction the diffusion coefficient function produces a map of diffusion coefficients B. The anisotropy arises from the independent application of the coefficient function in each direction. The gradients are multiplied by the diffusion coefficients and summed to reconstitute the total change in all directions. The time step λ controls the

rate of diffusion and must be sufficiently small (typically $\frac{1}{4}$ or $\frac{1}{7}$) to ensure numerical stability. At each step the diffused image from the previous step is the input for the next step.

$$\textbf{diffuse}(I, B) = I + \lambda(B_N * \nabla I_N + B_S * \nabla I_S + B_E * \nabla I_E + B_W * \nabla I_W) \qquad (9)$$

The system contains two free parameters which are discussed in the following section. The first, K, has already been described. Perona and Malik [33] propose a method for setting this parameter using Canny's noise estimator [12]. This and other adaptations of diffusion coefficients are discussed in the Diffusion Coefficients section below. The stopping time determines how many iterations of diffusion are performed. Perona and Malik do not provide an automatic method to set this parameter and opt for hand-picked constants instead.

Although it is a very effective and popular image regularisation technique, several problems were found with the original anisotropic diffusion formulation. You et al. [50] identify some of these including stair-stepping, ill-posedness, and numerical instability and perform a significant mathematical analysis of these behaviours. Stair-stepping is the tendency of anisotropic diffusion to introduce sharp edges in regions where the intensities change gradually. The ill-posedness is because two similar but noisy images can diverge under anisotropic diffusion instead of converging to more similar smoothed forms. Numerical instability is typically an implementation problem solved by using a smaller time step λ for the finite differences scheme. Two years later Weickert [47] claims that the mathematical foundations for the ill or well-posedness of anisotropic diffusion are inconclusive and the only obvious negative artefact of the algorithm is stair-stepping. Several adaptations and alternate formulations have been proposed, but many researchers, including this author, continue to use the original equations.

3.2 Diffusion Coefficients

Perona and Malik recommend the Canny noise estimator [12] to set the scale constant: "a histogram of the absolute values of the gradient throughout the image was computed, and K was set equal to the 90% value of its integral"[33]. The absolute value of the gradients in each direction are combined in an integral histogram and K is chosen such that the 10% strongest edges are preserved. In most of their examples, however, Perona and Malik use hand-picked scale constants. While effective in preserving image structure, the original coefficient with this criteria for selecting the scale constant has a tendency to soften edges, perform badly under sharp noise, and introduce artefacts in regions with gradual intensity changes [50].

Black et al. [8] propose a solution to soft edges by reposing diffusion as an estimation problem where the diffusion coefficient is the norm. They show the reason for these soft edges is that the Perona and Malik norm never reaches zero to completely inhibit

diffusion. Using robust statistics the authors are able to clamp off diffusion entirely for identified edges. The edges appear as outliers in the author's statistical model based on Tukey's biweight estimator using median absolute deviation (MAD). The new regularisation norm doesn't require the tuning of a diffusion constant and recovers sharper edges. Rifkah et al. notice that the Black coefficient cannot distinguish between edges and noise when the DC component of the noise is similar to the DC component of the edges. This type of noise is often manifested as sharp intensity spikes that don't form complete edges. The authors replace the MAD function the mean absolute deviation (MEAD) which is influenced more by this type of noise [36].

An alternate method for dealing with noise in the form of singular spikes is to calculate the diffusion coefficient over a support region. You et. al. mention that using a support region to isotropically diffuse the coefficients is an easy way to make the problem well- posed, but embedding an isotropic process in an anisotropic one "is obviously against the spirit of anisotropic diffusions" [50]. This can be accomplished with a simple Gaussian convolution of the input image before taking the gradients. To filter irregularities in optical flow fields Proesmans et. al. calculates the gradients using a support window instead of applying a Gaussian blur to the input [34]. Finally, to solve the same problem Yu et. al. replace finite differences with a kernel method that more reliably separates edges from noise [51]. While these adaptations solve problems related to rapid changes in image intensity, others have dealt with the stair-stepping phenomenon within gentle gradients.

Both You et al. and Weickert notice the stair stepping effect. Regions with gradual gradients cluster into constant-valued regions. The result is an appearance of thresholding with solid-coloured bands along the gradient strata. To solve the stair-stepping problem Hajiboli [21] proposes a fourth-order diffusion coefficient that incurs a significant additional computational expense. In practice stair-stepping rarely becomes a problem.

Finally, a method proposed by Scharr et al. learns the constant from a set of training data [37]. The relationship between the image statistics and the diffusion constant and stopping time is learned for a set of training data. This mapping is then used to diffuse previously unseen examples. The learning is highly specialised and a learned relationship is only valid for a specific class of images and single image geometry.

3.3 Stopping Criteria

When to halt diffusion is one of the most challenging parts of the original anisotropic diffusion equations; too little diffusion leaves residual noise while too much diffusion erodes the image structure. One approach is to rewrite the diffusion equations such that stopping time is a parameter that can be minimised [41]. Alternately, the Black et al. coefficient [8] separates edges and noise sufficiently enough that diffusion converges on the solution. In this case diffusion is stopped when an application of it has no

effect. Convergence can be induced by limiting how much each pixel can change from its original value [42]. Scharstein and Szeliski previously accomplished this isotropically with the membrane model [39]. Finally, both Scharr et al. [37] and Rifkah and Amer [36] learn the stopping time from a training set. Unfortunately the learning can only be extrapolated to similar images captured with the same imaging system.

3.4 Hierarchical Approaches

Although anisotropic diffusion is able to create an implicit scale space by varying its scale parameter, some work has been done on building explicit scale spaces with anisotropic diffusion. Acton introduced a full-octave image pyramid that is diffused at every level [3]. The pixels are iteratively linked to the most similar pixels in the neighbouring scales to create a segmentation and edge map [1]. Hierarchical anisotropic diffusion has also been formalised as a multigrid method for partial differential equations [2].

3.5 Anisotropic Diffusion for Disparity Refinement

Beyond the initial applications in image regularisation, edge detection, and segmentation, anisotropic diffusion has been used occasionally in optical flow and stereo vision. There is also one instance in the literature of it being used to determine shape from focus [18]. The optical flow community has primarily been interested in anisotropic diffusion for regularising the flow field and correcting the flow field at occluding boundaries. Typically anisotropic diffusion will be applied as a regularisation step after the flow field or disparities have been calculated [34, 39, 5]. It can be used to fill in occlusions [34] or completely integrated into the matching process [24]. To integrate anisotropic diffusion into the matching process, Ince and Konrad [24] propose a new cost function that relaxes the photometric constraint[1] if the geometric constraint[2] is violated. The cost function instead minimises at a neighbouring non-occluded pixel so long as the match does not cross an image intensity edge. The result is a cost function that mimics the behaviour of anisotropic diffusion.

Fua develops a stereo matching algorithm that rejects any matches that might be erroneous [19]. The result is a sparse disparity map with many holes. Intensity-edge guided anisotropic diffusion is used to fill the known disparities values into the neighbouring holes without blurring depth boundaries. Maabar [30] uses intensity edge-guided anisotropic diffusion in a Laplacian of Gaussian image pyramid to smooth disparity maps in a post-processing step (Section 4). The certainty of the match can also be used as a guiding criteria for the stereo matching process, but Banno and Ikeuchi take this *Directed Anisotropic Diffusion* a step further and also smooth the depth map using the squared dot product of the surface normals in place of the coefficient function

[1] limits the amount matched pixels can differ in colour or intensity

[2] limits the amount any two adjacent disparities can differ to enforce smoothness

[5].

The disparity refinement technique proposed by Banno and Ikeuchi is most similar to the approach proposed in this dissertation (Section 4). Instead of using a continuous mapping between confidence and the diffusion coefficients they use hard thresholds and inhibit diffusion entirely in low-confidence areas (Equation 10). Diffusion occurs from pixel p into pixel q and each pixel has been label as having high confidence H or low confidence L. The diffusion coefficient then becomes $g'(\nabla I) = w_{pq} * g(\nabla I)$ where

$$w_{pq} = \begin{cases} 1 & \text{if } p \in H \\ 0.5 & \text{if } p, q \in L \\ 0 & \text{if } p \in L, q \in H \ . \end{cases} \tag{10}$$

The approach proposed in this dissertation differs from Banno and Ikeuchi's because it does not use discrete labelling and thresholding. It also diffuses the confidence map to evolve the confident regions and produce a stopping criteria.

4 Investigation of Anisotropic Diffusion for Disparity Refinement

4.1 Introduction

Two different approaches for disparity refinement are presented in this section. The first embeds the Maabar intensity edge-guided anisotropic diffusion process [30] into the C3D stereo matching algorithm. The second approach, CIGAD, combines the intensity and confidence information to produce an improved disparity refinement technique. In the final results CIGAD is compared to the closely related C3D and Banno and Ikeuchi methods. These three methods are summarised in Table 1. Before describing the refinement approaches the datasets are presented because the first technique depends on ground truth data to determine the stopping criteria.

Table 1: Comparison of the regularisation methods employed by C3D, CIGAD, and Banno and Ikeuchi [5]

	C3D	CIGAD	Banno and Ikeuchi
Disparity Smoothing	Weighted average	Anisotropic diffusion	Anisotropic diffusion
Weighting	Confidence	Intensity and confidence	Intensity and confidence
Confidence Type	Continuous confidence metric	Binary confidence labelling	Continuous confidence metric
Intensity Smoothing	N/A	Isotropic smoothing	Anisotropic diffusion
Confidence Smoothing	N/A	N/A	Anisotropic diffusion

4.2 Datasets and Ground Truth

The stopping criteria for anisotropic diffusion in the initial approach "Intensity Edge Guided Anisotropic Diffusion with RMSD Ground Truth" depends on the existence of ground truth data for the scene being matched. Reliable disparity measurements must be available for the scene being captured to measure the noise reduction achieved by the refinement technique. This comparison is made to calibrate the scale and stopping time parameters. Although three data sets are used in this dissertation, the calibration and RMSD algorithm are only evaluated on the first dataset.

The first dataset is a scene of a dummy head set against a solid background (Figure 18) that was collected by Maabar [30]. No absolute ground truth model is available for this scene so a high-quality match on noise-reduced input is substituted instead. The frame averaging technique captures and averages one hundred images of the static scene to eliminate variation from Shott noise, crosstalk, heat, and other sources of

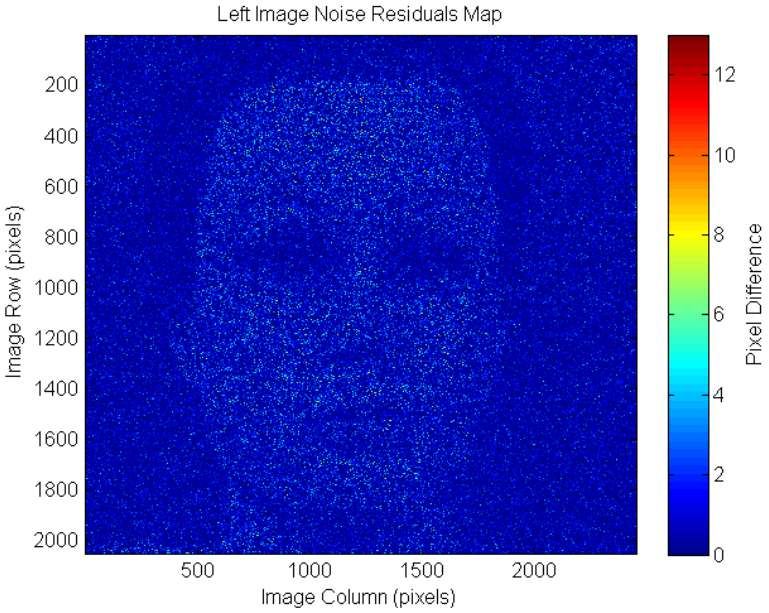


Figure 10: Left Image Noise Residuals Map

sensor noise. The frame averaged image A can be described as the mean of N sample images I_N (Equation 11) and the noise is measured by the sample standard deviation σ_A (Equation 12). The theoretical foundations for this type of regularisation was established by Nalcioglu and Cho [32]; the noise decreases asymptotically towards a lower bound as the number of samples increases. The rate of convergence depends on the dynamics of the specific imaging system, and the amount of noise suppression that can be achieved is bound by thermal noise. The random noise is difficult to discern by comparing the single and averaged frames, but a map of the residuals after the frames are subtracted (Figure 10) reveals the noise. The C3D matcher is then run on this averaged dataset with aggressive settings (50 semi-pyramid levels) to produce a refined disparity map. Snapshots are saved at every pyramid level, and these snapshots are the ground truth for their corresponding level in the experimental match. Even though robust matches can be produced using this method, it is impractical for real-world use. First, it is unusual to be able to capture 100 frames of a still human or animal, and this is the reason the experiment is performed on a dummy head instead of a real face. Second, the long execution time for the matching algorithm using 50 pyramid levels in unsuitable for robotics applications. This ground truth can, however, be used to calibrate the diffusion parameters for live operation.

$$A = \frac{1}{N}\sum_{i=1}^{N} I_i \tag{11}$$

$$\sigma_A^2 = \frac{1}{N}\sum_{i=1}^{N}(I_i - A)^2 \tag{12}$$

The second dataset is from the widely used Middlebury evaluation. The Middlebury

dataset is comprised of four scenes: Tsukuba, Venus, Teddy, and Cones. The Tsukuba image (Figure 24) first appears in a publication from the Nakamura et al. study on occlusions [31]. The scene was captured from many viewpoints in a planar grid and the ground truth disparities were hand-labeled. The Venus image (Figure 29) appears in Scharstein and Szeliski's original evaluation work [40]. The ground truth is determined by translating a camera horizontally and tracking features on the surfaces of the planar objects. The disparities of the features are extrapolated to the entire surface. The Teddy and Cones scenes (Figures 34 and 39) appear in a later paper by Scharstein and Szeliski [38] that uses structured light to obtain high quality disparity maps for the ground truth. Over one hundred matching algorithms are compared using these four scenes at `http://vision.middlebury.edu/stereo/`.

The third dataset also appears in the Middlebury collection but not as part of the standard benchmark. The "Cloth 1-4" set of images (Figures 44, 49, 54, and 59) give four different scenes of textured cloth at seven horizontal displacements, three light levels, and three exposure lengths. Horizontal displacements 1 and 5 are chosen to maintain a displacement consistent the benchmark images. The middle values for both illumination and exposure provide typical lighting and image quality. The ground truth data is captured using the same structured light method as [38].

4.3 Intensity Edge Guided Anisotropic Diffusion with RMSD Ground Truth

Intensity Guided Anisotropic Diffusion refers to the technique developed by Maabar [30] where the intensity image gradients are substituted for the disparity gradients in the diffusion coefficients. The goal is to inhibit diffusion in the disparity map across areas that contain edges in the intensity image based on the assumption that depth edges tend to occur at intensity edges. Intensity edge guided anisotropic diffusion and C3D seem to complement each other; Maabar's method [30] decomposes the disparity map into a LoG pyramid before the regularisation is performed, and the C3D matcher decomposes the input images into a DoG pyramid for hierarchical warp matching. Instead of applying to diffusion to each level of a LoG pyramid, diffusion is applied to the disparity map after each level match of the DoG semi-pyramid.

$$\mathbf{rmsd}(x_1, x_2) = \sqrt{\frac{\sum_{i=1}^{n}(x_{1,i} - x_{2,i})^2}{n}} \tag{13}$$

The root mean squared difference (RMSD) between the ground truth and the matched disparities is used to optimise two of the anisotropic diffusion parameters: scale k and number of iterations c. The time step λ is fixed at 0.25. RMSD is minimised with a brute force search over fixed domains of k and c. The values of k vary between 10^-4 and 4 according to Table 6. For each of these values of k diffusion is run for $c = 256$ steps, the RMSD is calculated after each iteration, and the minimum is stored. For time efficiency and to remove outliers caused by the image background,

only a region of interest centred on the eyes, nose, and face is considered. Also, if for any iterations of a trial of k the RMSD increases diffusion is halted and the previous value is taken for the minimum. Algorithm 2 performs these steps for a single semi-pyramid level. The tuple $\{k, c\}$ returned by AD_RMSD minimises the RMSD at that scale in the image pyramid. For each level of the pyramid the pair is stored to complete the training step of the algorithm. Other scenes can then be matched using these same settings given that the scene is similar and the frame geometry is the same. These limitations are similar to the constraints of other learning approaches [37, 36].

Algorithm 2 Intensity Edge Guided Anisotropic Diffusion with RMSD Ground Truth

D: Disparity Map
D_{GC}: Ground Truth Disparity Map
I: Intensity Image
K: Array of diffusion constants to test
c: Maximum number of diffusion iterations
B: Diffusion Coefficient Map
MIN_{rmsd}: Array of RMSD for each K
MIN_c: Array of c values that produce the respective MIN_{rmsd}
min$(array)$: Returns the index of the minimum value set in $array$

```
1: procedure AD_RMSD(D, D_GC, I, K, c)
2:     for i ← 1 : sizeK do
3:         B ← g_pm2(∇I, K[i])
4:         MIN_rmsd[i] ← 1000
5:         for j ← 1 : c do
6:             D_temp ← diffuse(D, B)
7:             cur_rmsd ← rmsd(D_GC, D_temp)
8:             if cur_rmsd < MIN_rmsd[i] then
9:                 MIN_rmsd[i] ← cur_rmsd
10:                MIN_c[i] ← j
11:                D ← D_temp
12:            else
13:                break
14:            end if
15:        end for
16:    end for
17:    min_index = min(MIN_rmsd)
18:    return{K[min_index], MIN_c[min_index], D}
19: end procedure
```

4.4 Confidence and Intensity Edge Guided Diffusion

The second method uses a weighted average between a novel confidence coefficient and the $\boldsymbol{g_{pm2}}$ coefficient for the intensity image. The confidence coefficient g_c (Equation 14) is based on a sigmoid transfer function that maps the range of confidence gradients to approximately $[0, 1]$. The curve is scaled by the hardness constant H to favour positive confidence gradients causing high confidence areas to diffuse into low confidence areas. The value Y, computed at every diffusion step, is the maximum absolute value of

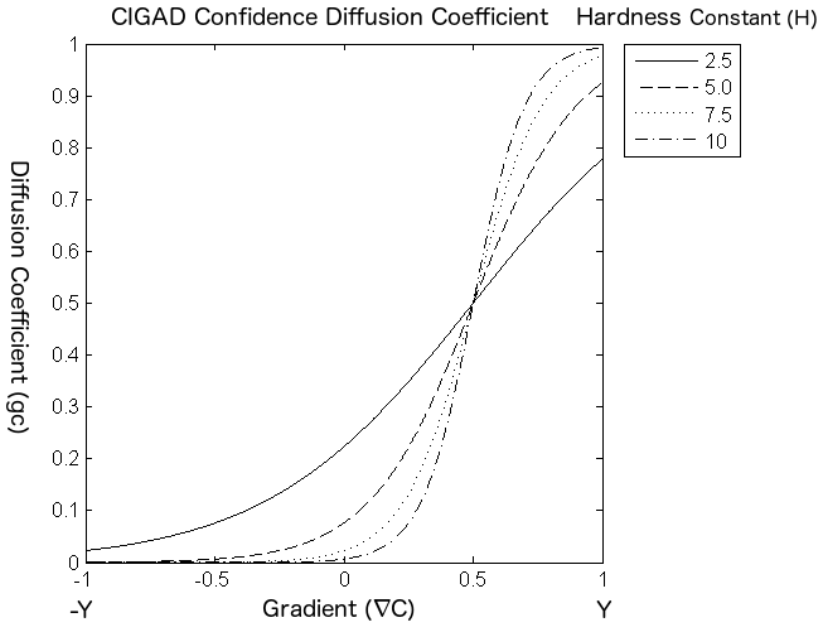


Figure 11: CIGAD Confidence Coefficient, Varied Hardness (Equation 14).

the confidence gradients. For Figure 11 the value of Y_c is set to 1. A value must be chosen that minimises the area under the curve (Figure 11) over $[-Y_c, 0]$ while still inhibiting diffusion at small positive gradients to prevent disparity map feature erosion. After evaluating plots of the function while varying H (Figure 11) a value of 10 was chosen for all of the experiments. The chosen value brings the output of g_c close to zero for zero gradients to inhibit diffusion where confidence does not change. Similarly, $H = 10$ causes g_c to output 1 for gradients near Y_c to increase diffusion where the confidence decreases. The k_2 term in Equation 14 adjusts the equation to different images' dynamic ranges. While it seems intuitive to replace the maximum with the mean or median for increased robustness this is not the case. These robust statistics provide an overly conservative gradient level estimate and that results in excess noise remaining in the disparity map. While a metric based on the maximum confidence gradient would normally be susceptible to outliers, the confidence map is diffused over many iterations and k_2 is calculated after each iteration; any outliers are quickly smoothed over and their effects are minimal.

$$k_1 = \frac{-H}{\nabla C_{max}}$$
$$k_2 = \frac{\nabla C_{max}}{2}$$
$$\mathbf{g_c}(\nabla C) = \frac{1}{1 + e^{k_1 * (\nabla C - k_2)}} \tag{14}$$

The second diffusion coefficient proposed by Perona and Malik (Equation 8) is applied to the intensity gradients ∇I. Through empirical study it was determined that setting the diffusion constant K to the mean of the gradients provides an appropriate

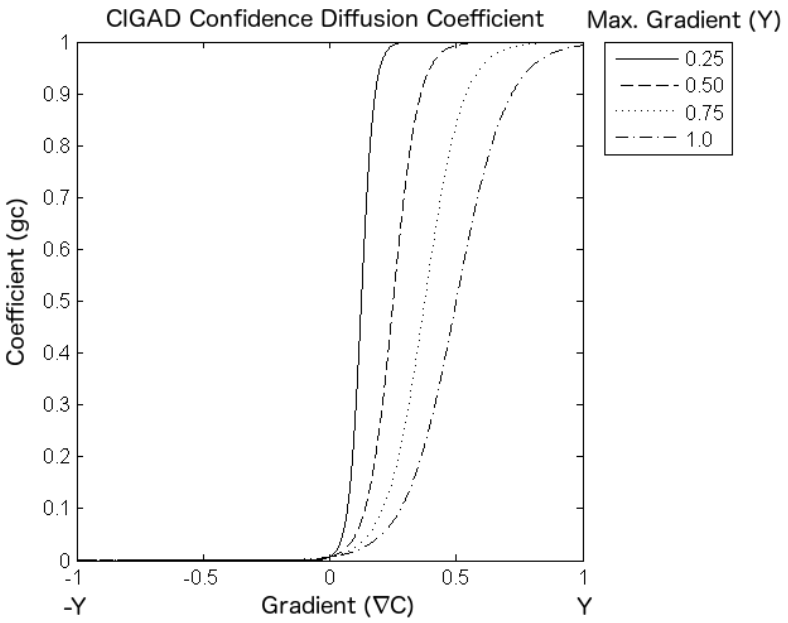


Figure 12: CIGAD Confidence Coefficient, Varied Max. Gradient (Equation 14).

amount of inhibition at the intensity edges. A sample of partial disparity maps from pyramid semi-levels were chosen, and the diffusion constant for each was hand-tuned. The hand-tuned values was always near the mean of the gradients. Calculating K based on the mean gradient eliminates the need to calibrate the method using ground truth such as in Algorithm 2 and adapts the algorithm to the varying matching scales. The two coefficient functions are combined in a weighted average with constant w to then diffuse the disparity map (Equation 15). Figure 13 plots the function along its two degrees of freedom: intensity gradient and confidence gradient. The maximum intensity gradient Y_I is calculated in the same manner as Y_c. The amount of diffusion is at its maximum in regions of decreasing confidence and constant intensity. Diffusion is nearly halted along high intensity gradients where the confidence increases. The ridge in the middle permits some diffusion to occur in all constant intensity regions to smooth noise in these areas.

$$\boldsymbol{g_{cigad}(I, C)} = (1 - w) * \mathbf{g_c}(\nabla C) + w * \boldsymbol{g_{pm2}}(\nabla I, \overline{|\nabla I|}) \tag{15}$$

The CIGAD algorithm (Algorithm 3) proceeds in two stages. First the intensity image is diffused for a fixed number of iterations with the $\boldsymbol{g_{pm2}}$ diffusion coefficient and using the mean gradient $\overline{\nabla I}$ for the scale constant. This step removes noise from the intensity image and provide adaptive support to the image gradients as opposed to the Gaussian convolution as suggested by You et al. [50] that provides uniform support. Pixels that receive a greater amount of diffusion have a wide area of support for their gradient calculation while pixels with inhibited diffusion (edges) have an appropriately small area of support. Since the intensity coefficient is reused to diffuse the disparity map this also prevents noise from the intensity image from being introduced into the disparity map. Noise in the intensity image creates false edges that inhibit diffusion

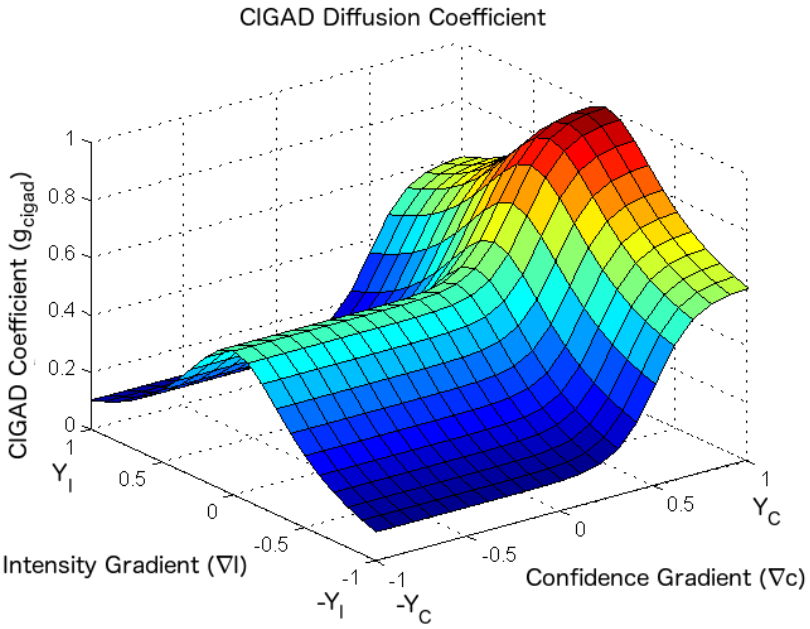


Figure 13: CIGAD Diffusion Coefficient, $w = 0.5$ (Equation 15)

in the disparity map. These random points of inhibition create artefacts manifested as noise in the disparity map. Next, in a loop that terminates after the confidence map converges or after a fixed number of iterations, the confidence map and the disparity map are diffused.

The stopping criteria depends on the $C_{MAD} > 0.001$ term of the while loop condition of Algorithm 3. The Mean Absolute Difference (MAD) for images (Equation 16) calculates the average difference between each pixel in two images of the same dimension. The MAD of two consecutively diffused confidence maps can determine if diffusion has reached sufficient convergence. The magnitudes of pixel changes are relatively large in the first few diffusion steps and begin to decrease as the system reaches equilibrium. The MAD gives a global estimate for these changes in magnitude and likewise asymptotically approaches zero with each iteration of diffusion. Therefore, a small MAD between two successively diffused confidence maps indicates the approximate convergence point for the diffusion system.

$$\mathbf{MAD}(X, Y) = \frac{1}{n * m} \sum_{i,j=0}^{n,m} |X_{i,j} - Y_{i,j}| \tag{16}$$

The visual effect produced by this diffusion process on the confidence map is that of a grass fire; the high confidence areas are expanded along their perimeter and they begin to merge. Diffusion in the disparity map occurs primarily along these perimeters because $\boldsymbol{g_{cigad}}$ responds positively to the large confidence gradients that occur in these areas. This wave is the effect that propagates information from high confidence areas into low confidence areas. Figure 14 demonstrates this effect on 100x100 synthetic images. The synthetic data is not a random dot stereogram but a completely synthetic disparity map, intensity image, and confidence map. The disparity map includes noisy,

Algorithm 3 CIGAD

MAD: Mean Absolute Difference

procedure CIGAD(D, C, I)
  for $j \leftarrow 1 : 10$ **do**
    $I_B \leftarrow \boldsymbol{g_{pm2}}(\overline{\nabla I})$
    $I \leftarrow$ **diffuse**(I, I_B)
  end for
  $i \leftarrow 1$
  $C_{MAD} \leftarrow 1$
  while $i \leq 50$ **and** $C_{MAD} > 0.001$ **do**
    $C_B \leftarrow \boldsymbol{g_c}(\nabla C)$
    $C_0 \leftarrow C$
    $C \leftarrow$ **diffuse**(C, C_B)
    $C_{MAD} \leftarrow$ **MAD**(C, C_0)
    $D_B \leftarrow \boldsymbol{g_{cigad}}(I, C)$
    $D \leftarrow$ **diffuse**(D, D_B)
    $i \leftarrow i + 1$
  end while
  return$D$
end procedure

low-confidence matches at intensity boundary regions. Iteration 0 represents the initial values for the process. The intensity image is highly-textured except for a 10 pixel wide ring with a uniform texture. The disparities for the two regions separated by this border are 1 for the outside and 0 for the inside. Within the border the disparities are noisy. The confidence is high everywhere except for within the texture-less border. As diffusion proceeds the noisy disparities are filtered and filled by the high confidence regions bordering them, and at 100 iterations there exists a well-defined boundary between the inner and outer regions. Although the process is run for 100 steps on synthetic images, the next section demonstrates how it takes far fewer steps when embedded in a matcher and run on real images.

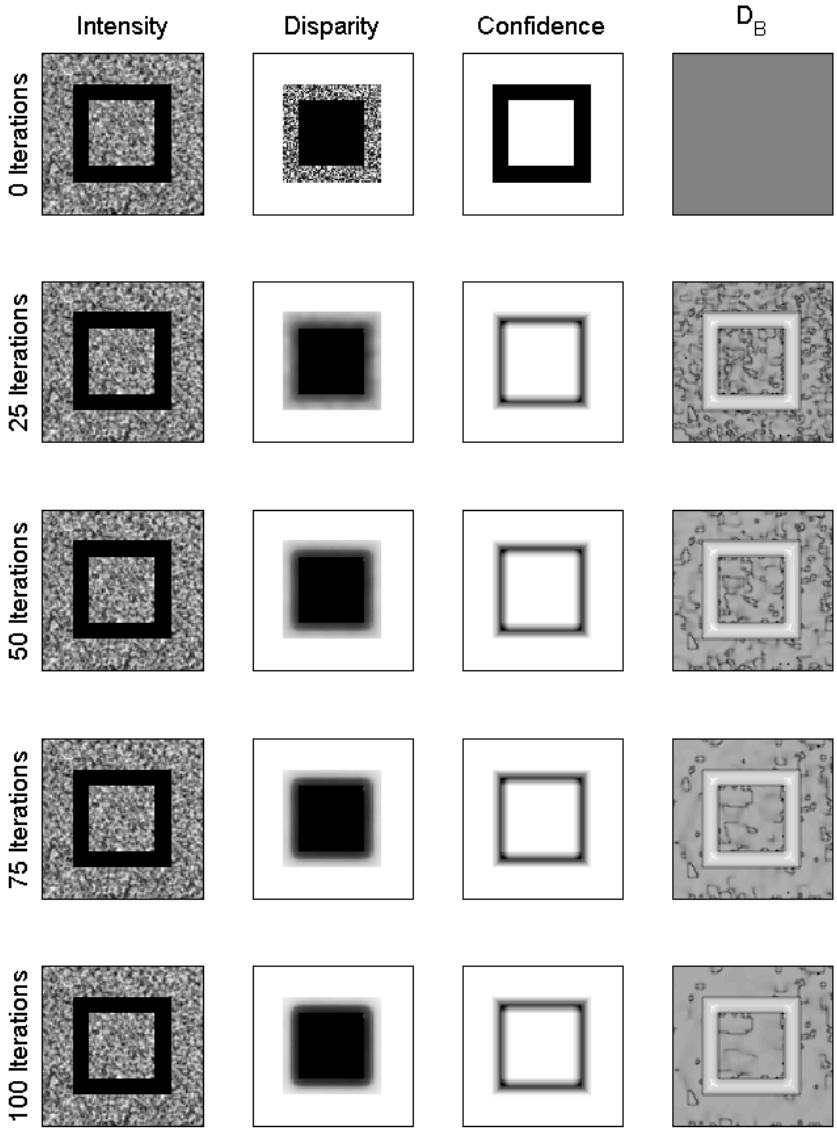


Figure 14: CIGAD on Synthetic Data

5 Experimental Results of Anisotropic Diffusion for Disparity Refinement

5.1 Introduction

The two different regularisation approaches, RMSD and CIGAD, are run on three different datasets, Mannequin Head, Middlebury stereo evaluation, and Middlebury Cloth. Both the standard C3D algorithm and the disparity refinement technique developed by Banno and Ikeuchi [5] provide baseline comparisons. The disparity refinement technique proposed by Banno and Ikeuchi is applied to the C3D matcher in the same way as the CIGAD algorithm. The left-right consistency check and diffusion step occur after every match iteration. The same parameters specified in [5] are used, and hand-tuning confirms that these parameters produce the best results for this method. The numerical evaluation includes only the Middlebury datasets because the Mannequin Head lacks a proper ground truth.

The Mannequin Head dataset comprises two crops of the same image. The first is the full frame with a solid background and the second is a cropped image of the facial features. The Middlebury stereo evaluation dataset comprises four images: Tsukuba, Venus, Teddy, and Cones. This dataset is commonly used to evaluate and compare stereo matching techniques. The four Middlebury Cloth images are also used in this dissertation. While these images do not commonly appear in the stereo matching literature, the cloth scenes are indicative of the CIGAD algorithm's potential for robot cloth folding applications. Thumbnails of the input images and all of the results figures appear in Appendix A to make this section more readable. The evaluation is similar to that of Scharstein and Szeliski [40]. The percentage of disparity errors are reported where an error is a difference from ground truth greater than one. Only "good" pixels are counted towards the score. For the Mannequin Face dataset a bit mask (Figure 15) is applied that excludes the textureless background. The ground truth provided by the Middlebury dataset gives special values to occluded pixels, and these occluded pixels are excluded from the score. The Middlebury dataset images are also evaluated based on their scaled mean disparity error. For each image the disparities are scaled by the width of the image and the mean error is determined. The combined means and standard deviations of all the images are used to perform a null hypothesis test.

5.2 Intensity Edge Guided Anisotropic Diffusion with RMSD Ground Truth

The RMSD method depends on a ground truth image for each level of scale space the matcher uses. There are two ways to generate this ground truth. The first method decomposes the full-resolution ground truth disparity map into a Gaussian semi-pyramid. Note that unlike an intensity image, when downsampling a disparity map the dispari-

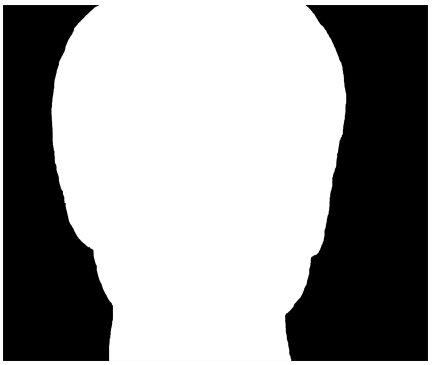


Figure 15: Mannequin Head Bitmask

ties must be divided by the downsampling factor. The second method takes disparity map snapshots from the C3D matcher at each level in the semi-pyramid; after line 10 of Algorithm 1 a copy of the disparity map, tagged with the semi-pyramid level, is written to disk to be re-used as ground truth at the corresponding semi-pyramid level in the RMSD algorithm. Each method has its pros and cons but neither was able to achieve better performance than baseline.

The first ground truth method provides the most complete disparity map against which to calculate RMSD at each matching step. Unfortunately, this ground truth method performs poorly in practice. The author hypothesises that the complete disparity map does not capture the state of the matcher during each step of the iterative process. C3D matches difference of Gaussians bandpassed images which each contain a limited spectrum of the spatial frequencies of the full resolution image; large features are matched first and smaller features are added on as the pipeline moves up through scale space. Except for the finest scale-space levels, the Gaussian pyramid ground truth images contain more disparity information than the matcher has calculated. A similar problem occurs when using disparity map snapshots. Even though the same matcher is used to calculate the ground truth and perform the experiments, the regularisation process greatly influences how the matching proceeds. In either case, the ground truth differs from the matcher to such a degree that diffusion is unable to reduce the RMSD between the two disparity maps. Both types of ground truth cause the RMSD to be consistently large, but disparity map snapshots provide the best performance. The large RMSD drives the scale constant k to its upper limit in Algorithm 2. Once k grows larger than the intensity image gradients, anisotropic diffusion becomes isotropic diffusion.

Evaluation of the RMSD method is only performed on the Mannequin head dataset because acquiring ground truth data in the necessary format is impossible for the Middlebury images. Even though the noise level is extremely reduced, the disparity map produced is shallow with eroded features (Figures 18c and 21c). The number of matching errors in Table 4 are much higher than the baseline C3D. Moderate to high amounts of error accumulate on the surfaces with larger gradients (Figures 19c and 22c) and relative to baseline, accuracy decreases across most of the face. The RMSD

produced fewer low-magnitude errors but more large-magnitude errors than both of the other methods (Figures 20 and 23). The RMSD method also causes the maximum error of 34.54 within the cropped region.

The bad performance of this method led to the hypothesis that a diffusion algorithm with more adaptive and selective parameters was needed to improve the matching performance. Furthermore, the use of ground truth to train the system proved cumbersome, and neither the Gaussian pyramid nor the disparity map snapshots are similar enough to obtain a meaningful RMSD estimate. The capture process is also time consuming and must be repeated whenever the scene lighting changes.

5.3 Confidence and Intensity Edge Guided Anisotropic Diffusion

The CIGAD approach is evaluated on both the Mannequin Head and the Middlebury datasets using the methods described in the Introduction. Three parameters must be set before running the algorithm. First, the diffusion time step $\lambda = 0.25$ prevents numerical instability while providing a fast rate of convergence, and this value appears frequently in the anisotropic diffusion literature. The weight w from Equation 15 that favours either inhibiting diffusion at edges or diffusing into low confidence areas is set to 0.5. Heavily favouring either the intensity-based coefficient (Equation 8) or the confidence-based coefficient (Equation 14) produces clearly worse results. Diffusion is halted when the MAD of two consecutive confidence maps is less than 0.001 or a maximum of 50 time steps are executed. Table 7 shows that, with the exception of the finest pyramid levels, diffusion is halted by the MAD criteria well before the 50 step maximum. Figure 16 demonstrates that when this maximum is reached the MAD is very close to its lower bound.

The mannequin face is evaluated using both the full frame and a cropped region over the centre of the face. Table 4 shows that in both cases CIGAD decreases the number of mismatches when compared the C3D and RMSD. The amount of noise is visibly less than C3D and the erosion caused by RMSD and Banno does not occur (Figure 18). The error is reduced at the edge between the face and the background except for the mannequin's left ear which is largely self-occluded (Figure 19). From the full-frame view the noise reduction over the face is not apparent, and it is for this reason the cropped version is included (Figure 22) where the noise reduction is obvious.

The evaluation is performed using background masks because the solid-coloured background is difficult to match and produces an unreliable ground truth estimate. Normally the matcher will produce nonsensical disparity measurements over a textureless region because it is matching noise in the image instead of texture features. CIGAD successfully suppresses this background noise as seen in Figure 17. The graceful degradation of the output is required to create a reliable range sensor for autonomous robot perception.

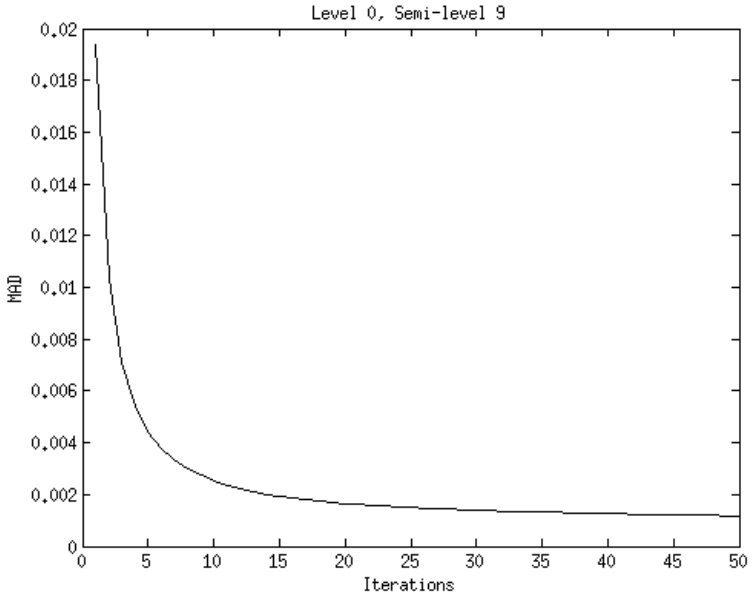
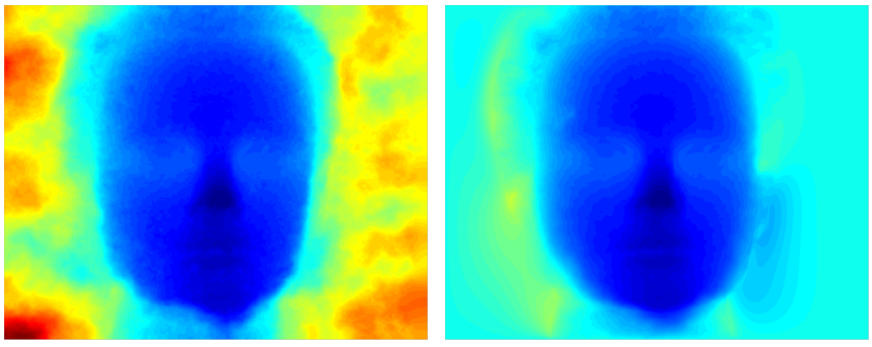


Figure 16: MAD vs. number of iterations for final matching step on the Mannequin Head dataset.



(a) C3D (b) CIGAD

Figure 17: Effective suppression of background noise

Next, each of the Middlebury stereo evaluation pairs are matched and the same evaluation is performed for each of the four images. The Banno and Ikeuchi method produces the fewest errors on the Tsukuba, Venus, and Teddy images while CIGAD yields the best result on Cones. First, both CIGAD and Banno improve the overall definition of the objects in the Tsukuba scene, but Banno achieves a greater degree of noise reduction over the planar regions of the disparity map. The most important improvement is CIGAD correctly matching of most of the lamp arm. In a robotics context, the omission of the arm from Figure 25b could result in a grasp failure or collision. The recovery of this element attests to the robustness of the new method. The merging of the two narrow rods into one large region is of little consequence for the intended robotics application. Next, in the planar Venus scene Banno again demonstrates superior noise reduction capabilities. The third scene, Teddy, presents the most difficulty for all techniques. The Banno approach produces the best results through greater noise reduction and an ability to recover the boundary between the white cloth and the background. The white edge between the two objects has little texture, and the two objects have similar greyscale intensity values. The lack of a clear intensity edge allows the disparity values from the textured background to bleed onto the foreground canvas, and in this case the LRC confidence method used by Banno is able to choose the correct disparity estimate. Finally, in the Cones scene neither C3D or CIGAD are able to recover the tips of the Cones, but CIGAD computes the most accurate result overall. The evaluation scores achieved by CIGAD are higher than the adapted Banno method and many of those listed on the Stereo Evaluation website. The CIGAD algorithm was not designed with this benchmark in mind, and it makes no assumptions about the shape of the objects in the scene. A large proportion of the area of the stereo evaluation scene contain piecewise-planar surfaces that are not important for the motivation of this dissertation. Banno and Ikeuchi [5] also note many of the top algorithms do not perform well on real-world data.

The Middlebury Cloth dataset confirms the CIGAD algorithm's suitability for vision in the robot cloth folding application. CIGAD outperforms the other methods on all but the simplest (Cloth 1) scene. While structurally simple with only a single sheet draped mostly flat on a surface, the Cloth 1 image provides little intensity variation along its only occluding boundary. CIGAD and Banno both depend on strong intensity edges to diffuse the disparity values correctly. Banno produces the worst result with a major match error at the top of the scene while CIGAD produces small errors along the occluding edges. The Cloth 2 scene consists of a pile of sheets with many folds and occlusions. Both CIGAD and Banno create large errors in the upper left corner which are less prominent in the C3D approach, but this error is less severe in the CIGAD method which also recovers more of the foreground. The Cloth 3 images feature one blanket laid on top of the other. All methods correctly match most of the scene, but Banno spreads the match error in the lower-left of the frame further than the other approaches. The last scene, Cloth 4, has the three sheets hanging vertically

Table 2: Null Hypothesis Test Criteria: The mean error of the baseline (C3D in Table 5) is subtracted from the mean error of each experimental result (CIGAD and Banno in Table 5) to compute each difference D. The mean ($\overline{D}$) and standard deviations (**s**) are calculated for each of these differences.

Dataset	CIGAD - C3D	Banno - C3D
Tsukuba	1.3535×10^{-05}	-1.2324×10^{-04}
Venus	-1.2991×10^{-04}	-4.5052×10^{-04}
Teddy	6.4059×10^{-04}	-1.2025×10^{-03}
Cones	-1.6447×10^{-04}	6.0054×10^{-05}
Cloth 1	1.3108×10^{-04}	2.2131×10^{-03}
Cloth 2	2.5537×10^{-03}	3.9453×10^{-03}
Cloth 3	-1.4374×10^{-04}	1.5515×10^{-03}
Cloth 4	5.7432×10^{-04}	-1.9074×10^{-04}
$\overline{D}$	4.3438×10^{-04}	7.2538×10^{-04}
s	9.1291×10^{-04}	1.7057×10^{-03}
2**s**	1.8258×10^{-03}	3.4114×10^{-03}

with horizontal steps on the left and right sheets. Although CIGAD produces the best overall result and recovers the vertical edges better than the other methods, its recovery of the horizontal step edges is worse. Overall the CIGAD method outperforms C3D and Banno for providing robust disparity (depth) measurements for the robot cloth manipulation application.

Finally, a null-hypothesis test on the scaled mean disparity errors gives a more conflicting view of the algorithms' performance. The horizontal disparity errors for each image and method are scaled by the width of the image before calculating the mean error for each image/method combination. Table 5 lists the mean error for each trial. The Banno and Ikeuchi method appears to achieve the lowest mean error on the largest number of images. To determine these improvements do not occur by chance, the null hypothesis states that the difference between the means is zero. The mean errors of the baseline method (C3D) are subtracted from the mean errors of each experimental method (CIGAD and Banno). The results are listed in Table 2; negative values indicate an improvement over the baseline while positive values indicate a decrease in performance. These differences are labeled D and their mean is $\overline{D}$. The standard deviation of D is calculated as s and twice the standard deviation of D is labeled $2s$. The means of neither experimental method are greater than the standard deviation, indicating that there is less than a 90% chance that the differences between the methods cannot be explained by random variation. Based on these results the null hypothesis cannot be rejected and the claim that either CIGAD or Banno improves on the baseline method cannot be made with great certainty.

5.4 Computational Cost

The author anticipated that a more directed diffusion algorithm could converge in a small number of steps. Table 7 indicates that while little diffusion is performed at

Table 3: An estimate of the number of floating point operations per level of the image semi-pyramid performed by the CIGAD (Algorithm 3) disparity refinement step.

Line	Operation	FLOP's (per pixel)	Iterations
3	$\overline{\nabla I}$	5	10
3	g_{pm2}	5	10
4	diffuse	9	10
9	∇C	4	[0, 50]
11	diffuse	9	[0, 50]
12	MAD	3	[0, 50]
13	g_{cigad}	13	[0, 50]
14	diffuse	9	[0, 50]
Total		57	280

coarser scales, the algorithm reaches its fixed maximum of 50 iterations at the finest scales. Since the algorithm performs the most iterations at the finest levels in scale space, the computational cost is considerable. Table 3 approximates the number of floating point operations (FLOP's) per line of the CIGAD algorithm. For n being the number of pixels in an image at a given scale, the worse case number of operations is 57 FLOP's per n pixels over 280 iterations or $15960 * n$. To refine the disparity map at the finest level of the image pyramid on a 5 megapixel image takes 7.98×10^{10} floating point operations. This number seems large but put into perspective of the performance of modern computers is not that high; Intel claims 14.4 GFLOP/s on their modest Core 2 Duo E4300 processor[3]. The entire stereo matching process on a 5 megapixel image typically takes about 5 minutes.

6 Conclusion and Future Work

The measurement of disparity from a stereo pair of images is imprecise in practice due to sensor noise and sparsely textured regions. A regularisation step can be included in the matching process to reduce the effect of these sources of error. This dissertation describes a new variation of anisotropic diffusion that uses confidence from the matching cost and the intensity image to diffuse high confidence disparity estimates into neighbouring low confidence regions while preserving depth discontinuities. It accomplishes this by diffusing the intensity image for adaptive gradient support, defining two new diffusion coefficients, and diffusing the confidence map alongside the disparity map. This new process is able to produce more reliable disparity maps than the baseline algorithm.

One observed deficiency in the CIGAD process is that visible intensity edges are not always strong enough to inhibit diffusion across a depth boundary. The human eye can perceive these differences but the changes in intensity are numerically small. Sometimes this is the case of a change in hue while the brightness remains the same. The

[3]http://download.intel.com/support/processors/core2duo/sb/core_E4000.pdf

matching and regularisation in this work was only performed in greyscale colour space, but another colour space could be chosen instead to increase the differences between contrasting colours. A similar problem occurs when the foreground and background are similar in colour but different in intensity texture. Texture segmentation could distinguish between these surfaces and augment the existing edge detection. Different confidence metrics could be considered, and different base matching algorithms could be evaluated as well. The results indicated that some approaches performed better under specific conditions. Multiple refinement algorithms could be run simultaneously with a voting scheme choosing the most accurate result.

A Disparity Refinement Results Figures

A.1 Overview

Table 4: Percentage of pixels with disparity errors greater than one. The **best scores** appear in boldface.

Dataset	C3D	CIGAD	Banno
Tsukuba	13.44	12.41	**12.25**
Venus	9.52	7.19	**3.69**
Teddy	26.11	27.03	**20.51**
Cones	23.53	**19.60**	21.65
Cloth 1	**10.09**	14.63	18.51
Cloth 2	24.84	**24.60**	34.50
Cloth 3	12.91	**11.93**	22.01
Cloth 4	25.76	**24.49**	29.49

Table 5: The mean error of every horizontal disparity map is determined by scaling the errors by the image width before computing the mean.

Dataset	C3D	CIGAD	Banno
Tsukuba	1.6062×10^{-03}	1.6197×10^{-03}	$\mathbf{1.4829}\times10^{-03}$
Venus	9.6707×10^{-04}	8.3716×10^{-04}	$\mathbf{5.1656}\times10^{-04}$
Teddy	3.2218×10^{-03}	3.8623×10^{-03}	$\mathbf{2.0192}\times10^{-03}$
Cones	3.1588×10^{-03}	$\mathbf{2.9943}\times10^{-03}$	3.2189×10^{-03}
Cloth 1	$\mathbf{7.7399}\times10^{-04}$	9.0507×10^{-04}	2.9871×10^{-03}
Cloth 2	$\mathbf{3.7966}\times10^{-03}$	6.3503×10^{-03}	7.7420×10^{-03}
Cloth 3	1.4267×10^{-03}	$\mathbf{1.2829}\times10^{-03}$	2.9782×10^{-03}
Cloth 4	6.4911×10^{-03}	7.0654×10^{-03}	$\mathbf{6.3003}\times10^{-03}$

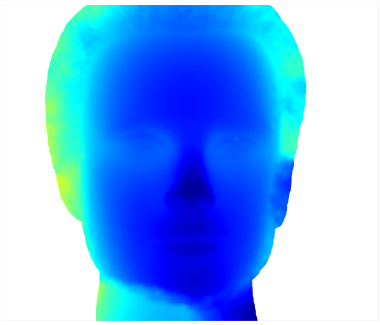
A.2 Mannequin Head

Table 6: Values of k used for the brute force search step of the RMSD approach.

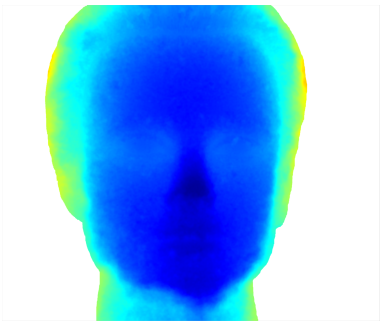
0.0001	0.0002	0.0003	0.0004	0.0005	0.0006	0.0007	0.0008	0.0009
0.0010	0.0020	0.0030	0.0040	0.0050	0.0060	0.0070	0.0080	0.0090
0.0100	0.0200	0.0300	0.0400	0.0500	0.0600	0.0700	0.0800	0.0900
0.1000	0.2000	0.3000	0.4000	0.5000	0.6000	0.7000	0.8000	0.9000
1.0000	1.5000	2.0000	2.5000	3.0000	3.5000	4.0000		

Table 7: Number of CIGAD iterations per pyramid level.

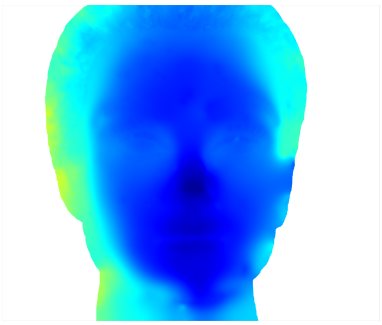
Pyramid Level \ Semi-pyramid Level	0	1	2	3	4	5	6	7	8	9
0	50	50	50	50	50	50	50	50	50	50
1	49	49	49	49	50	50	50	50	50	50
2	47	47	45	45	45	47	48	50	50	50
3	35	36	36	37	38	39	41	43	46	47
4	25	26	26	26	26	27	29	31	32	34
5	19	17	17	17	17	18	19	21	22	25
6	13	12	11	10	10	11	12	13	15	18
7	10	9	8	7	8	8	9	10	11	14
8	6	6	5	5	5	5	5	6	6	6
9	4	3	4	4	4	4	4	3	4	4
10	4	2	0	0	0	0	0	0	0	1
11	9	7	4	1	0	0	0	0	1	0
12	11	9	4	3	0	0	0	0	0	1
13	0	0	0	0	0	0	0	0	0	0
14	0	0	0	0	0	0	0	0	0	0
15	0	0	0	0	0	0	0	0	0	0



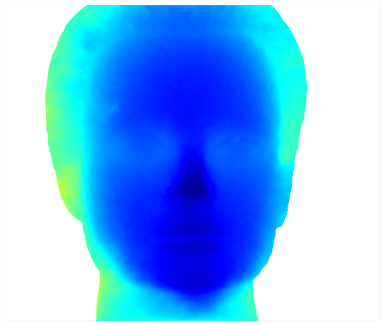
(a) Ground Truth



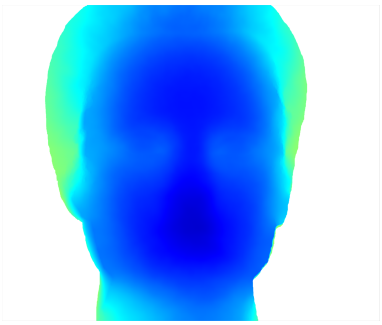
(b) C3D



(c) RMSD

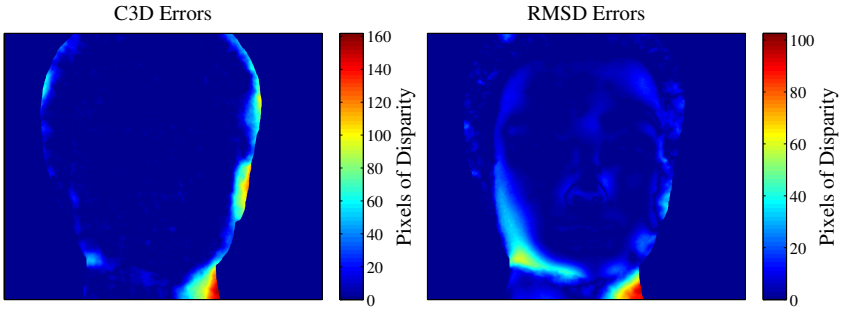


(d) CIGAD



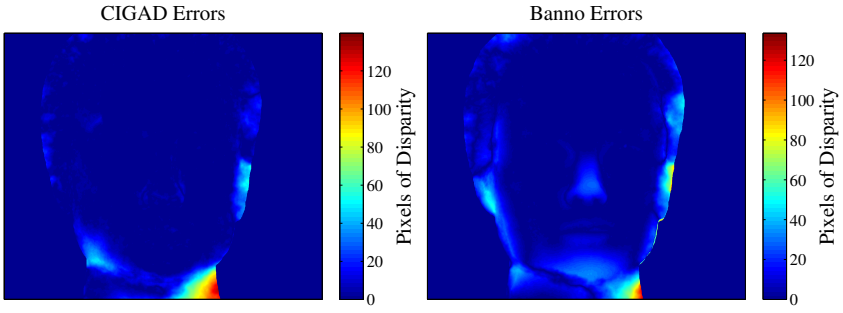
(e) Banno

Figure 18: Mannequin Head Disparity Maps



(a) C3D

(b) RMSD



(c) CIGAD

(d) Banno

Figure 19: Mannequin Head Disparity Errors

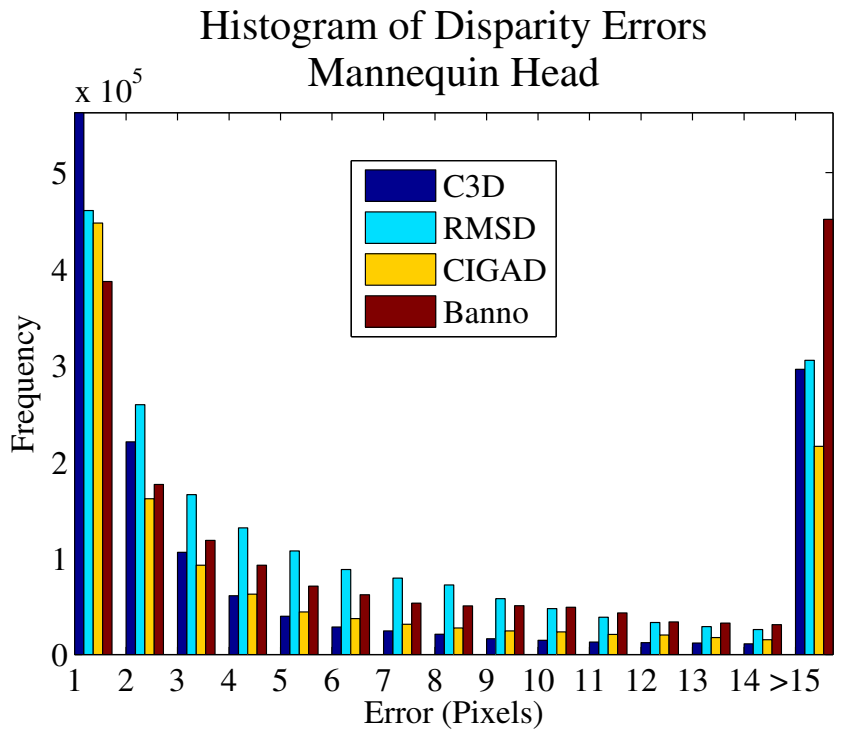
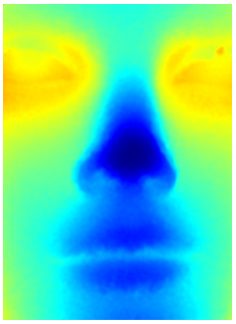
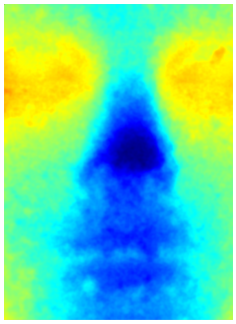


Figure 20: Mannequin Head Error Histogram

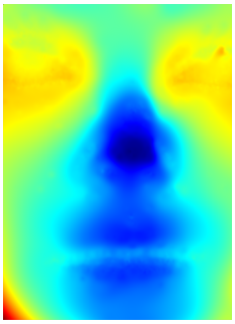
A.3 Mannequin Head Cropped



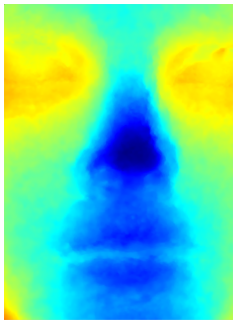
(a) Ground Truth



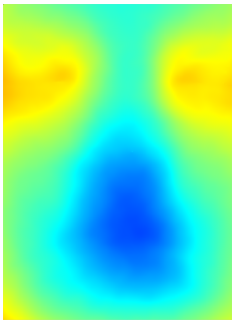
(b) C3D



(c) RMSD

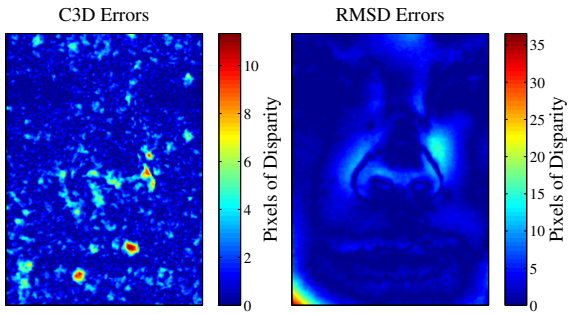


(d) CIGAD

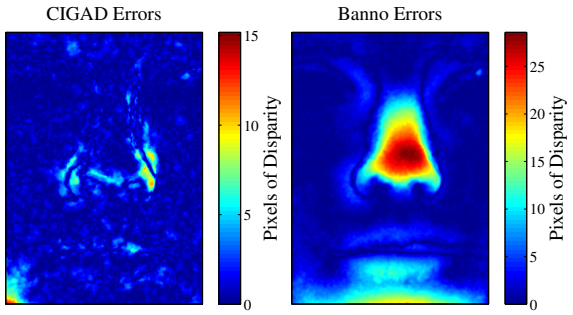


(e) Banno

Figure 21: Mannequin Head Cropped Disparity Maps



(a) C3D (b) RMSD



(c) CIGAD (d) Banno

Figure 22: Mannequin Head Cropped Disparity Errors

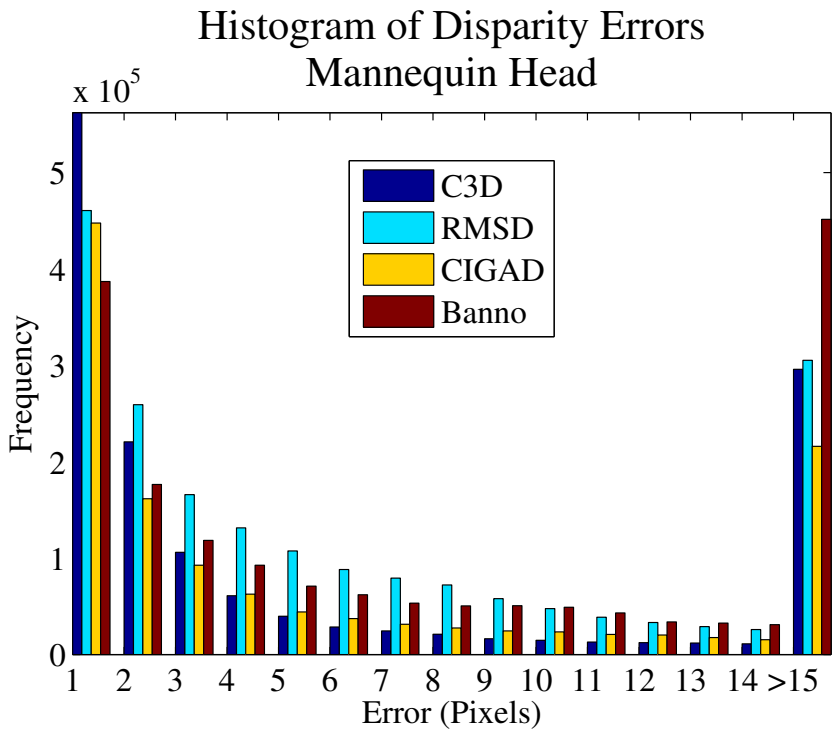
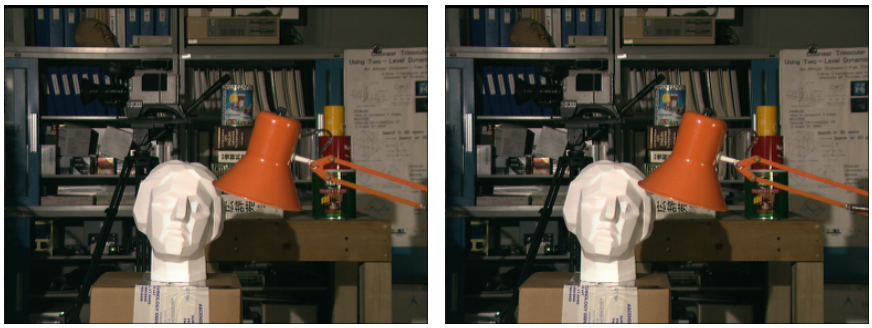


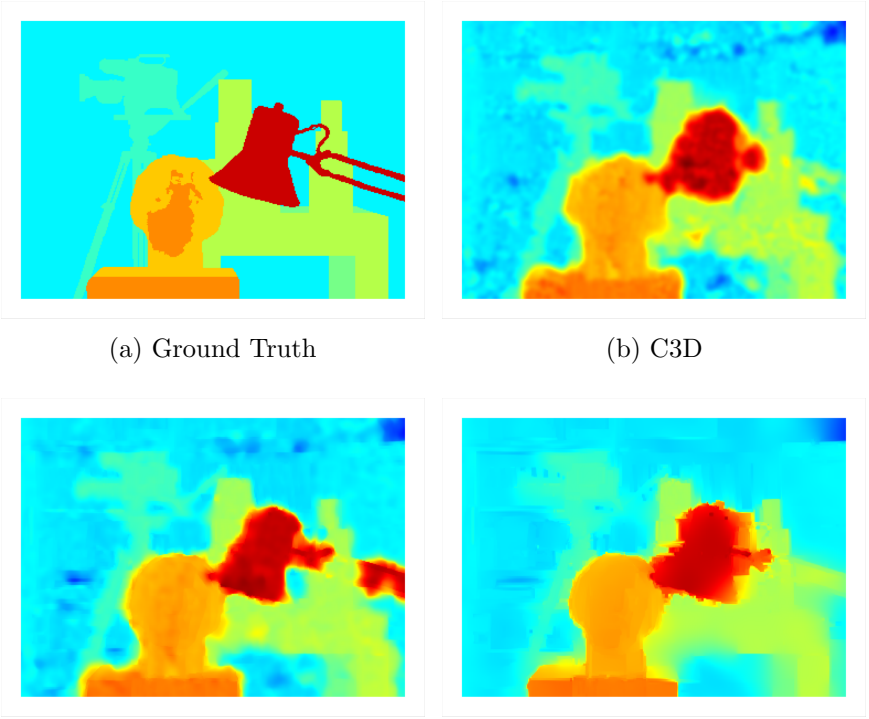
Figure 23: Mannequin Head Cropped Error Histogram

A.4 Tsukuba



(a) Left Frame (b) Right Frame

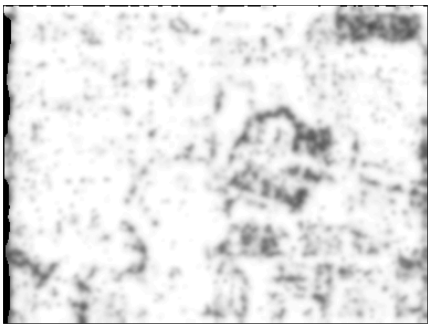
Figure 24: Tsukuba Input Frames



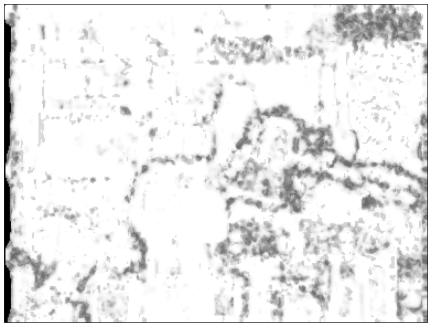
(a) Ground Truth (b) C3D

(c) CIGAD (d) Banno

Figure 25: Tsukuba Disparity Maps



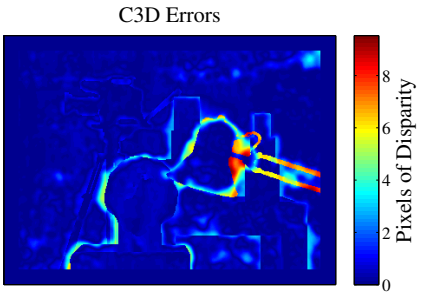
(a) C3D



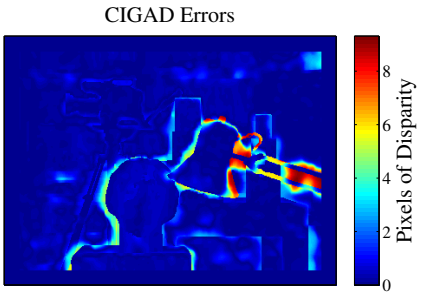
(b) CIGAD

(c) Banno

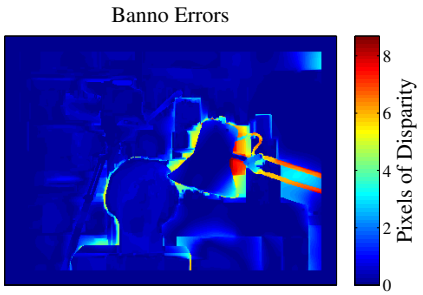
Figure 26: Tsukuba Confidence Maps. White indicates a high confidence while black indicates a low confidence.



(a) C3D



(b) CIGAD



(c) Banno

Figure 27: Tsukuba Disparity Errors

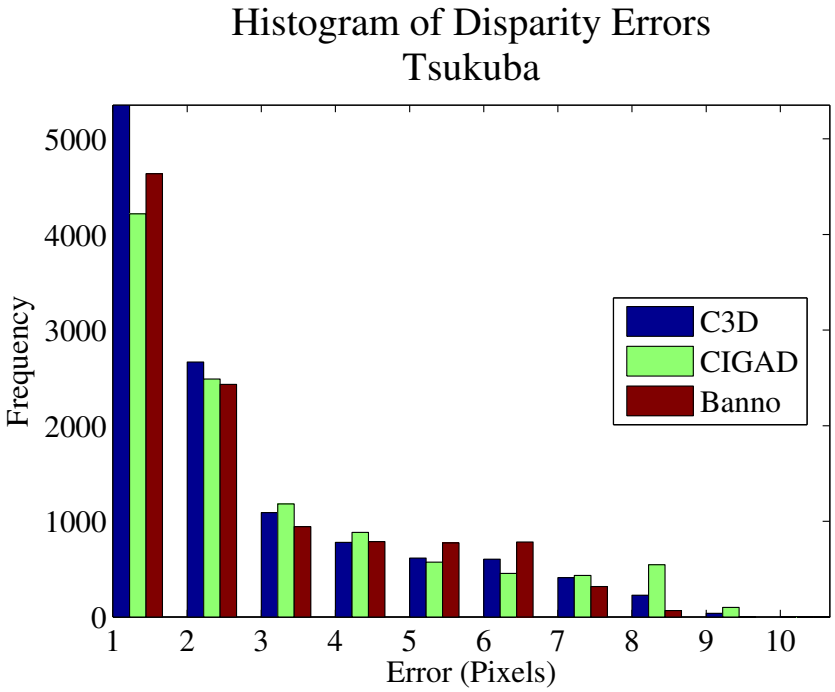
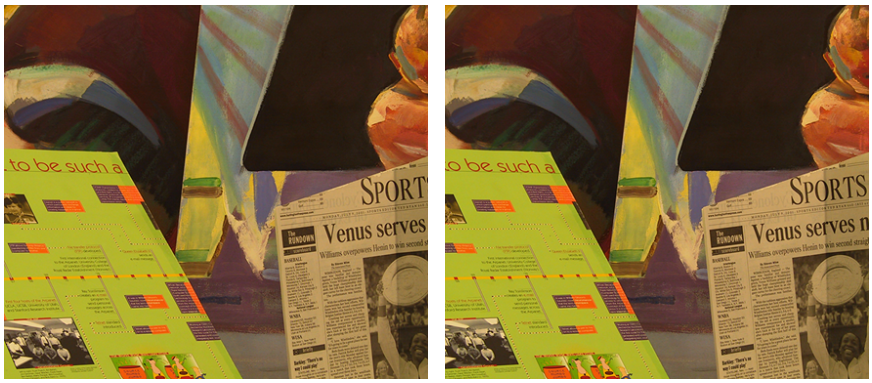


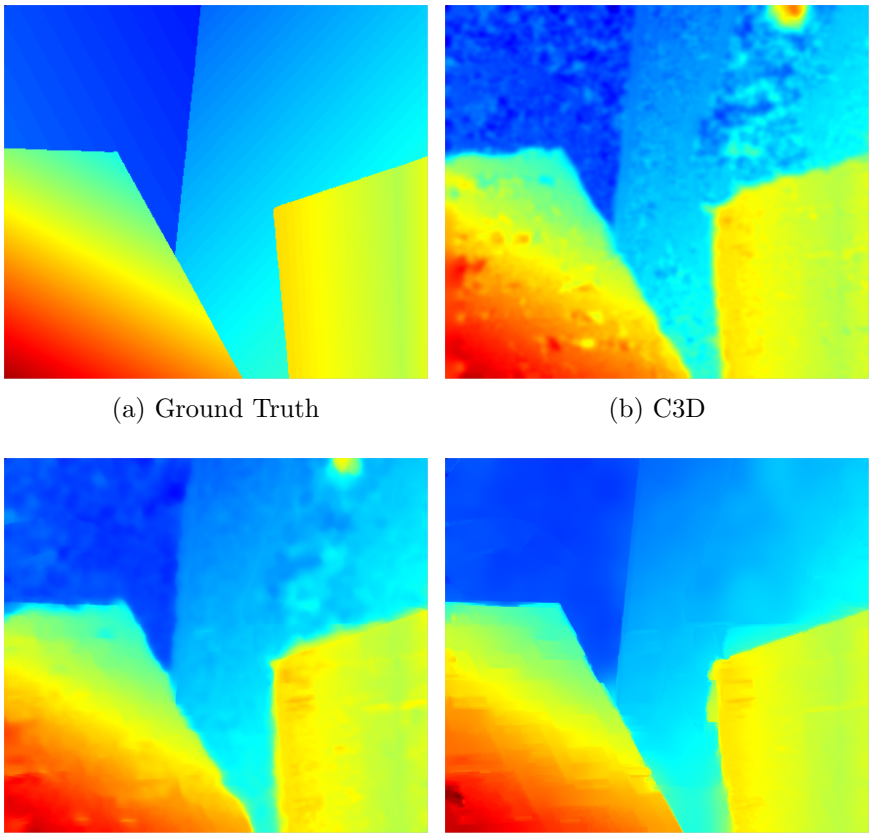
Figure 28: Tsukuba Error Histogram

A.5 Venus



(a) Left Frame (b) Right Frame

Figure 29: Venus Input Frames



(a) Ground Truth (b) C3D

(c) CIGAD (d) Banno

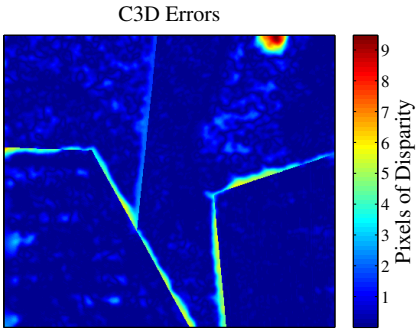
Figure 30: Venus Disparity Maps

(a) C3D

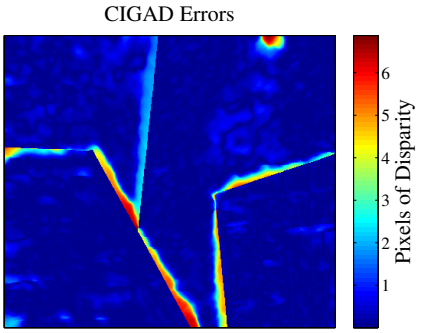
(b) CIGAD

(c) Banno

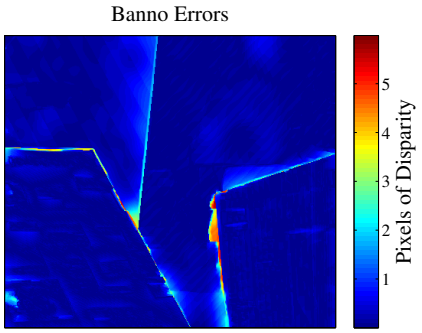
Figure 31: Venus Confidence Maps. White indicates a high confidence while black indicates a low confidence.



(a) C3D



(b) CIGAD



(c) Banno

Figure 32: Venus Disparity Errors

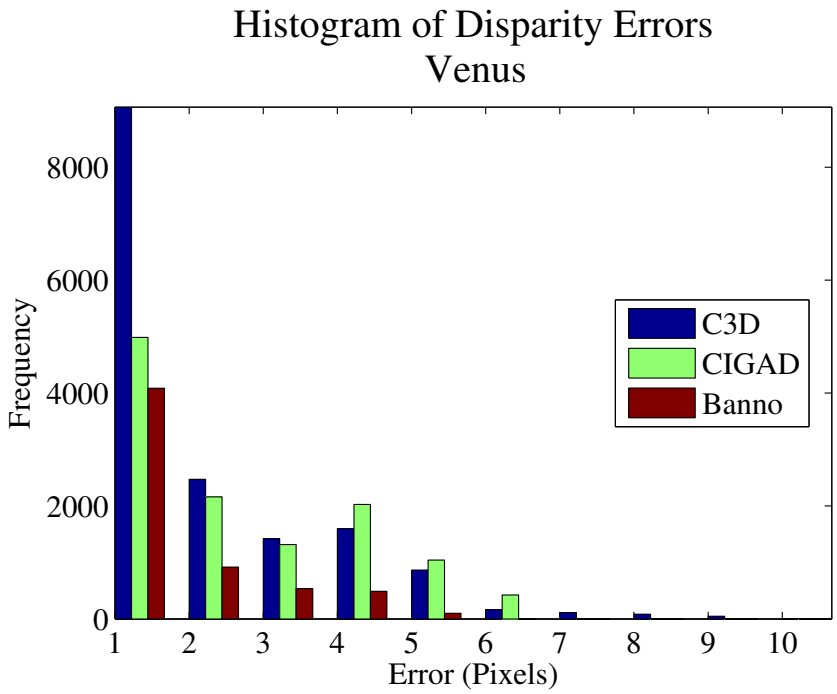
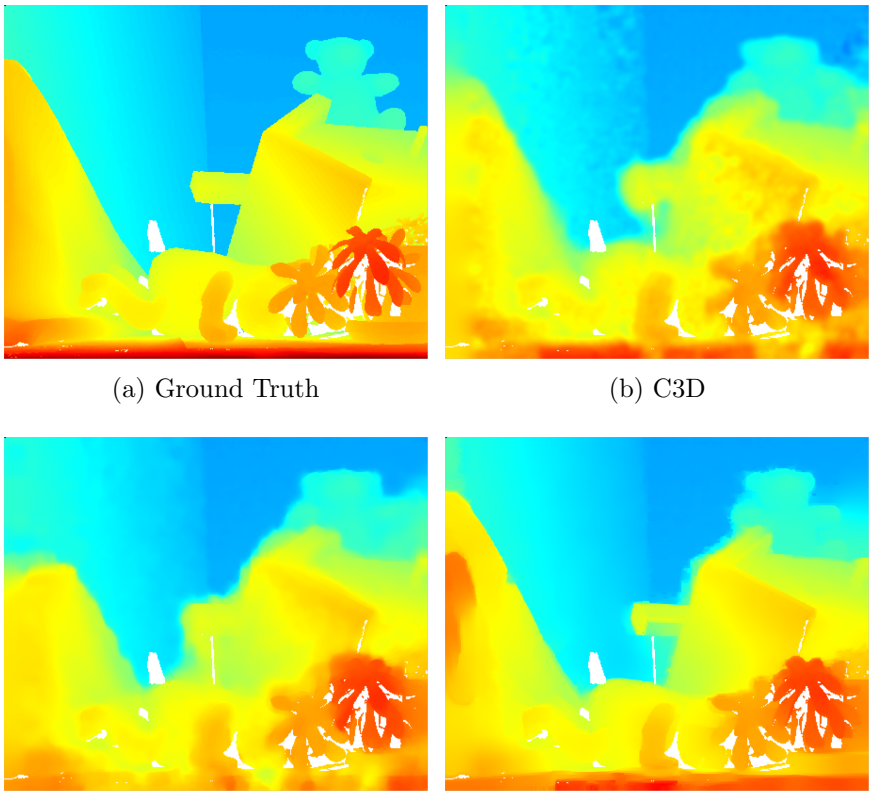


Figure 33: Venus Error Histogram

A.6 Teddy

(a) Left Frame (b) Right Frame

Figure 34: Teddy Input Frames

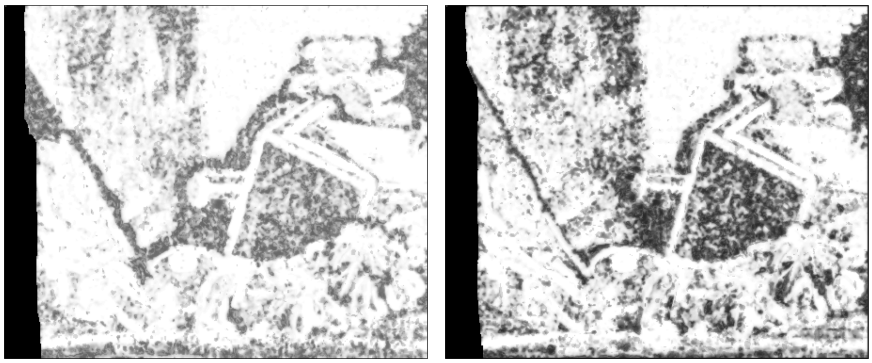


(a) Ground Truth (b) C3D

(c) CIGAD (d) Banno

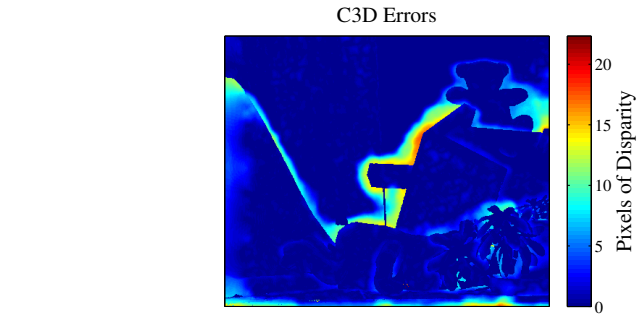
Figure 35: Teddy Disparity Maps

(a) C3D

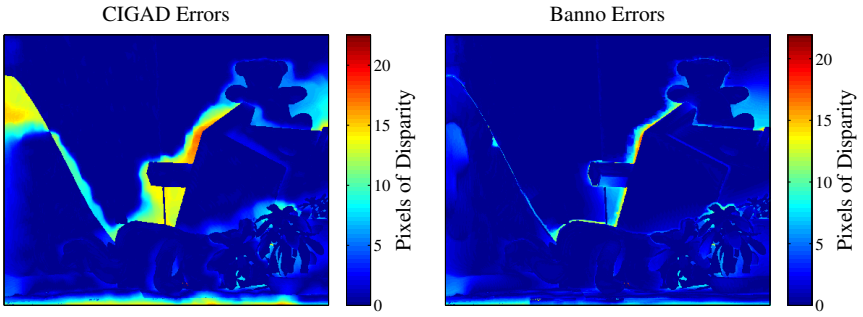


(b) CIGAD (c) Banno

Figure 36: Teddy Confidence Maps. White indicates a high confidence while black indicates a low confidence.



(a) C3D



(b) CIGAD (c) Banno

Figure 37: Teddy Disparity Errors

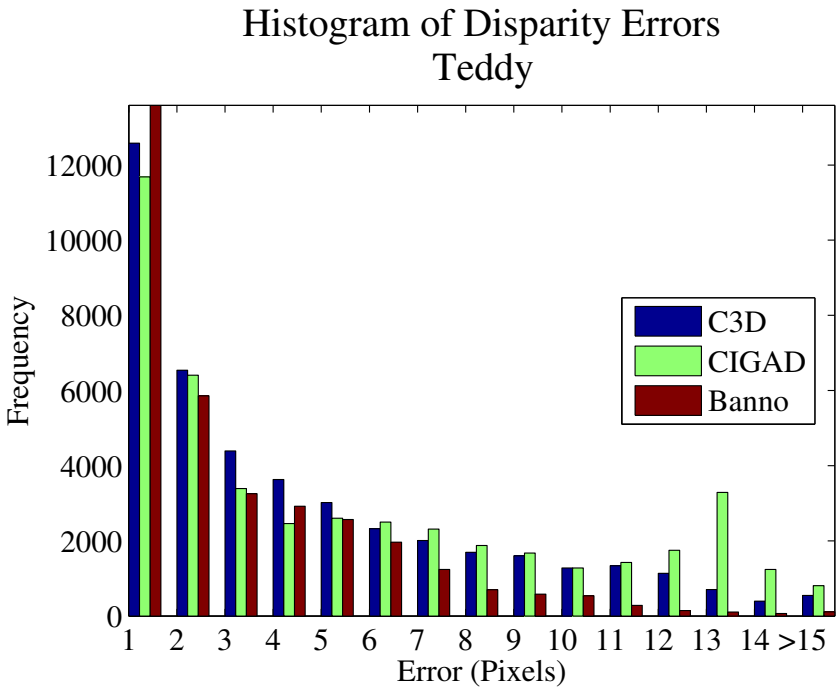
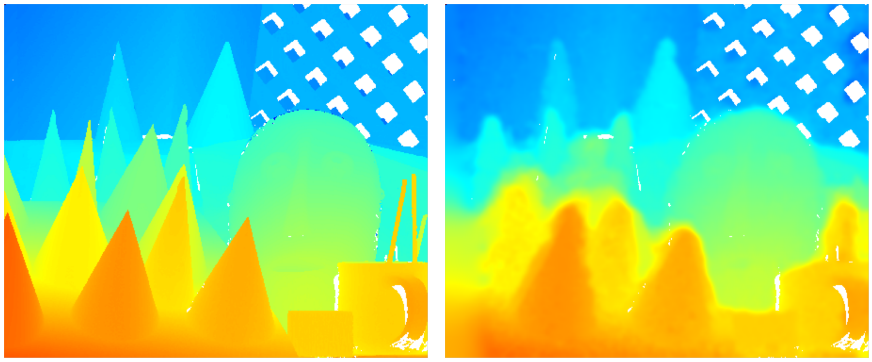


Figure 38: Teddy Error Histogram

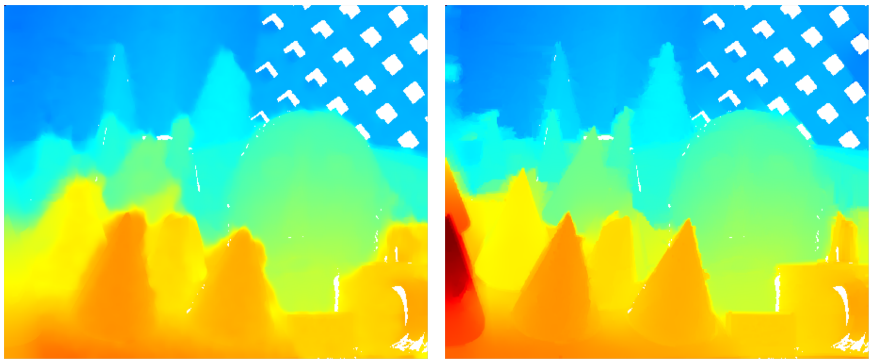
A.7 Cones

(a) Left Frame (b) Right Frame

Figure 39: Cones Input Frames

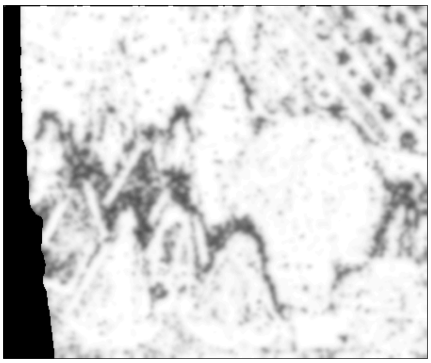


(a) Ground Truth (b) C3D

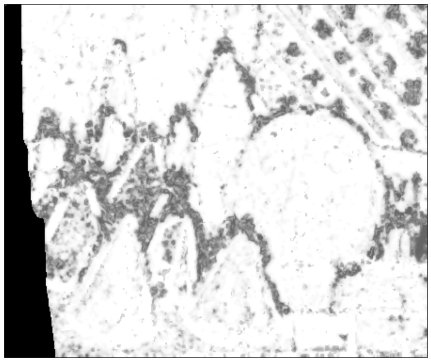


(c) CIGAD (d) Banno

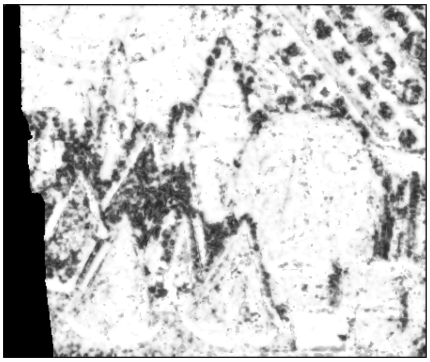
Figure 40: Cones Disparity Maps



(a) C3D

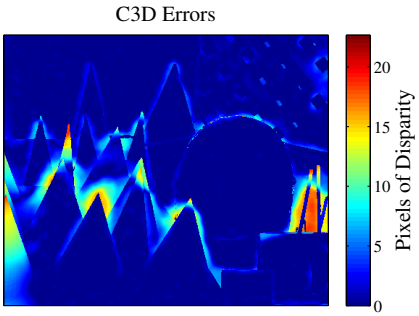


(b) CIGAD

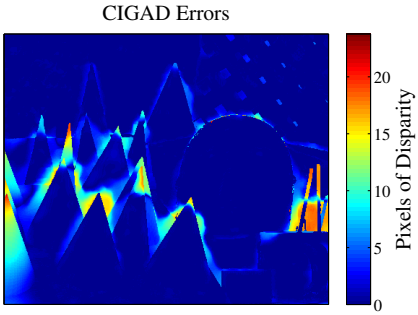


(c) Banno

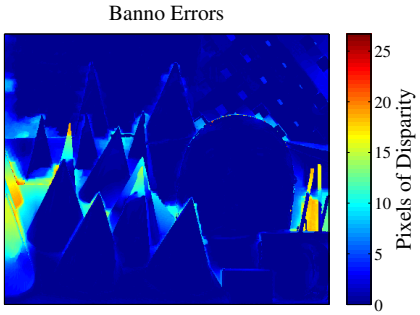
Figure 41: Cones Confidence Maps. White indicates a high confidence while black indicates a low confidence.



(a) C3D



(b) CIGAD



(c) Banno

Figure 42: Cones Disparity Errors

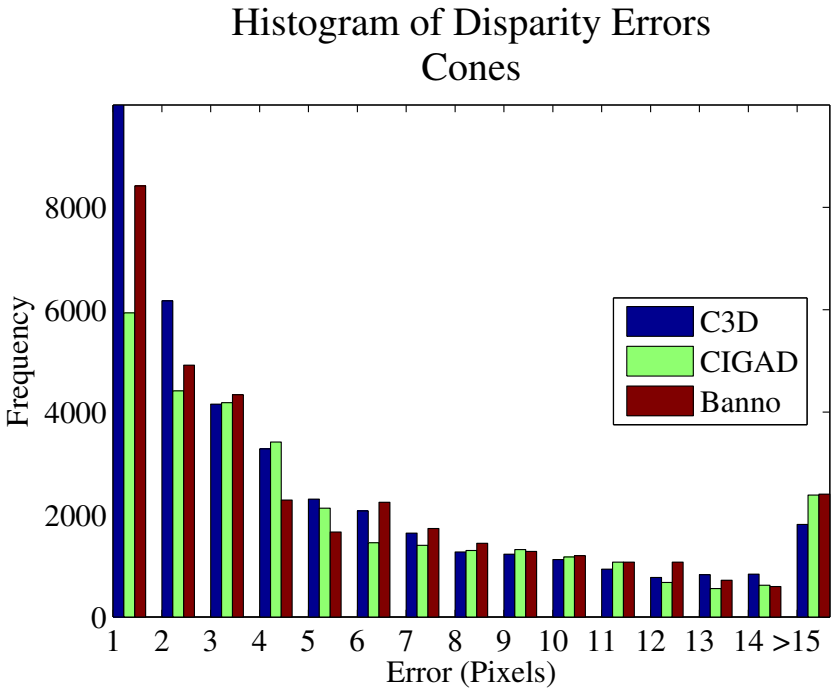
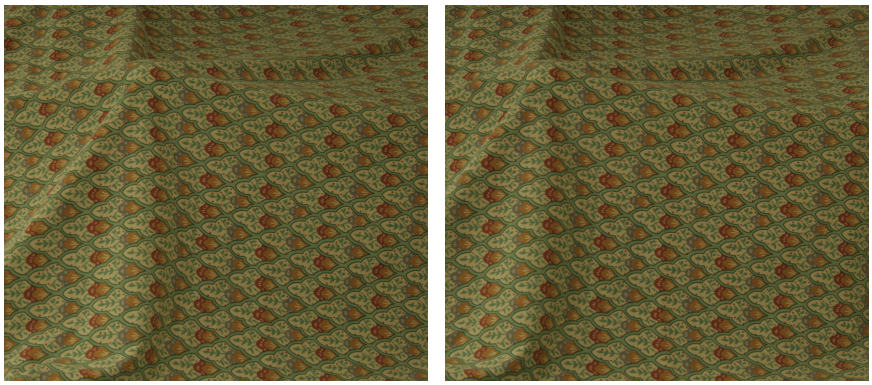


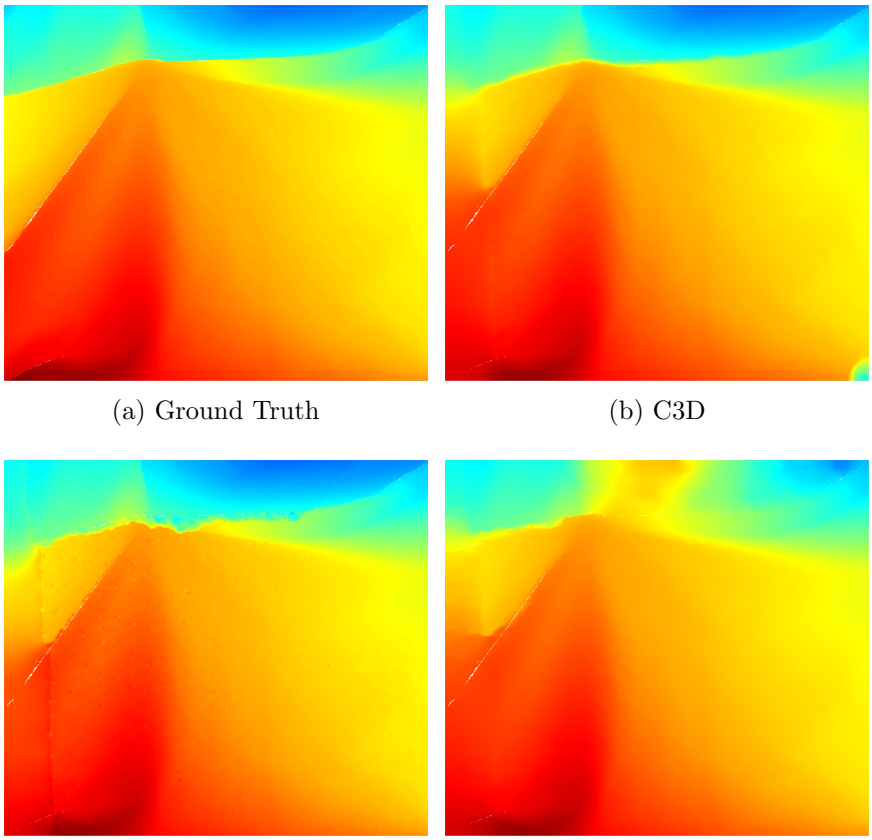
Figure 43: Cones Error Histogram

A.8 Cloth 1



(a) Left Frame (b) Right Frame

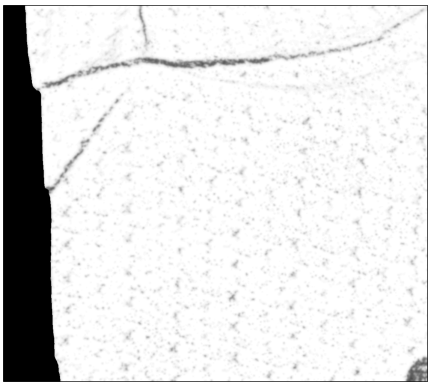
Figure 44: Cloth 1 Input Frames



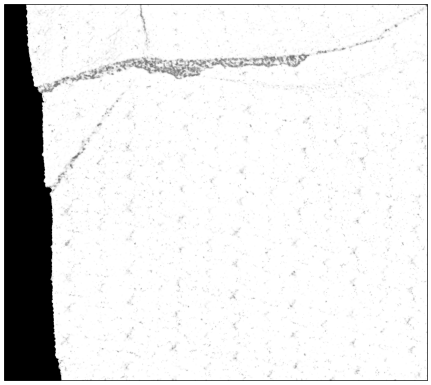
(a) Ground Truth (b) C3D

(c) CIGAD (d) Banno

Figure 45: Cloth 1 Disparity Maps



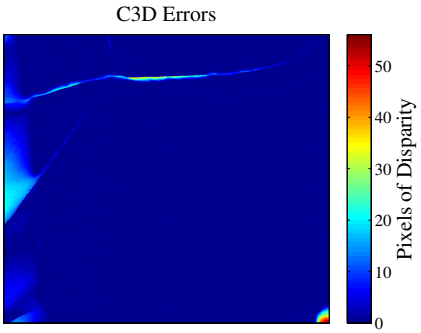
(a) C3D



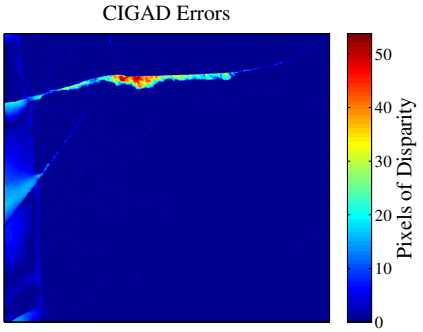
(b) CIGAD

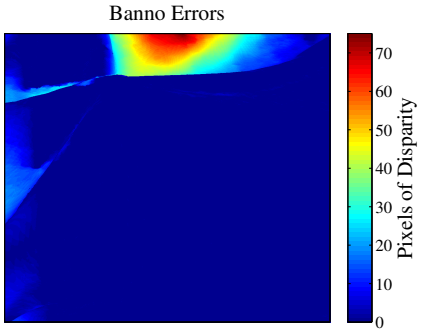
(c) Banno

Figure 46: Cloth 1 Confidence Maps. White indicates a high confidence while black indicates a low confidence.



(a) C3D





(b) CIGAD

(c) Banno

Figure 47: Cloth 1 Disparity Errors

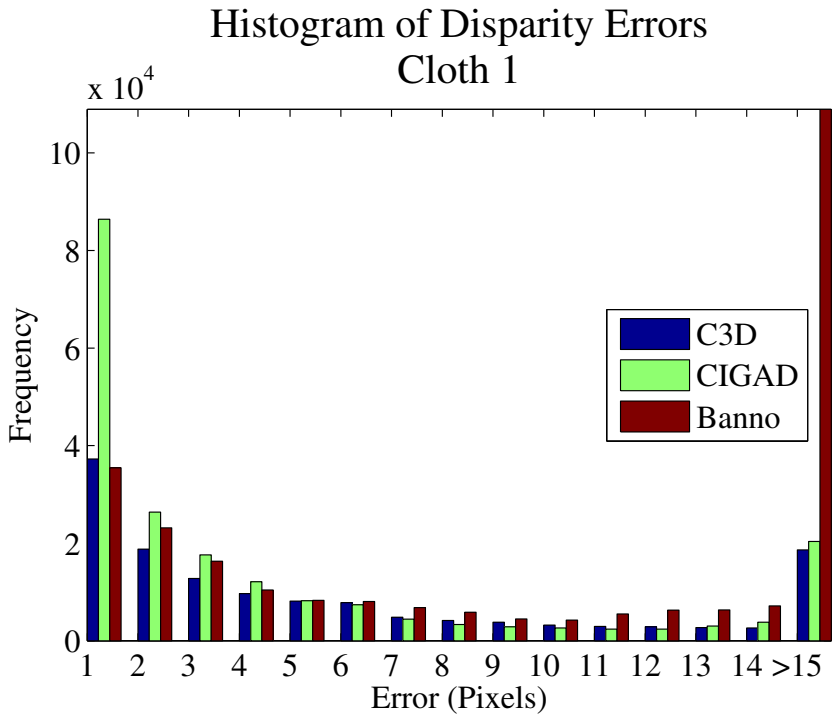
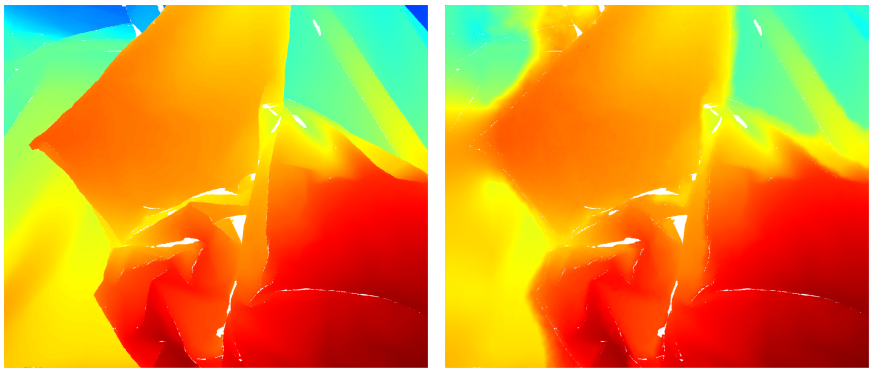


Figure 48: Cloth 1 Error Histogram

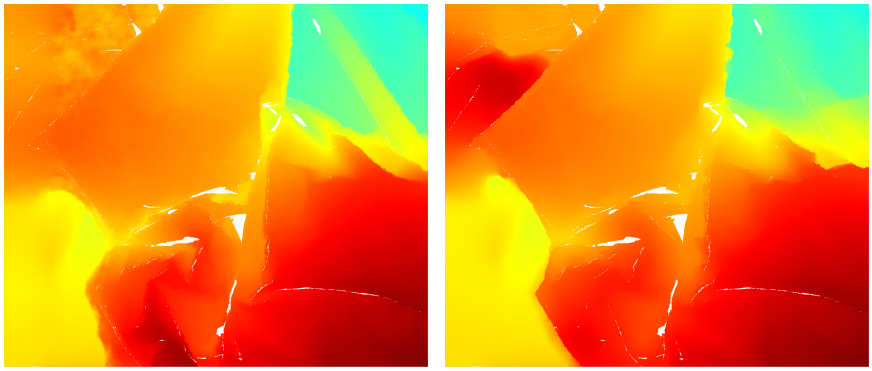
A.9 Cloth 2

(a) Left Frame (b) Right Frame

Figure 49: Cloth 2 Input Frames

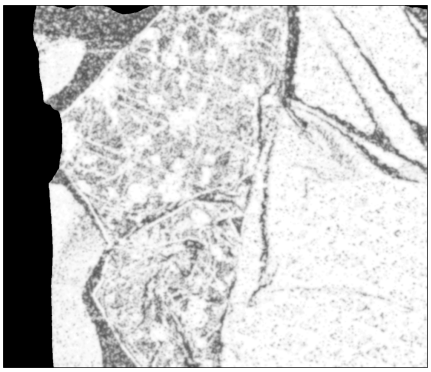


(a) Ground Truth (b) C3D

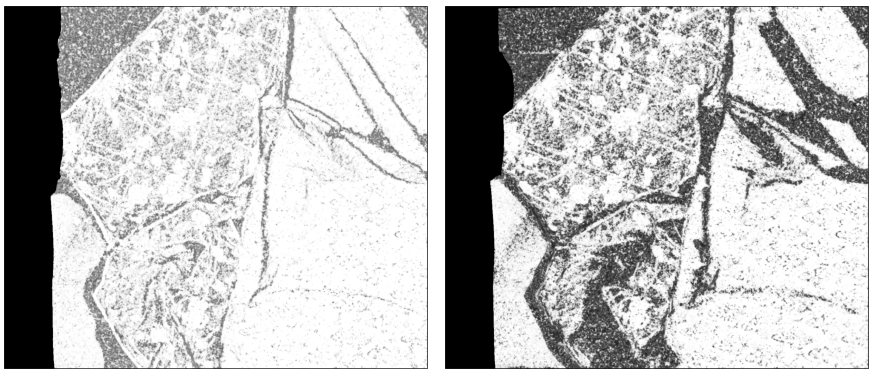


(c) CIGAD (d) Banno

Figure 50: Cloth 2 Disparity Maps

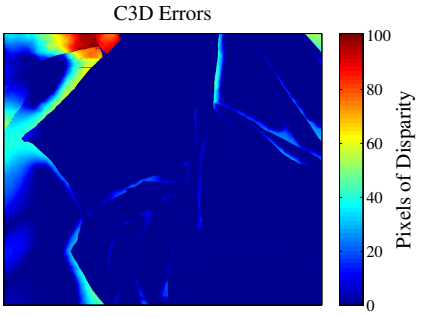


(a) C3D

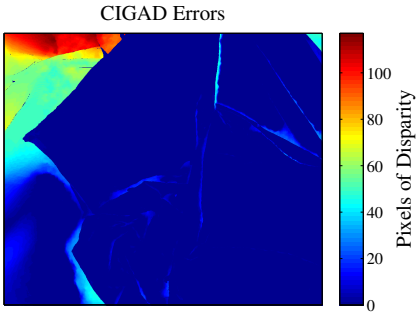


(b) CIGAD (c) Banno

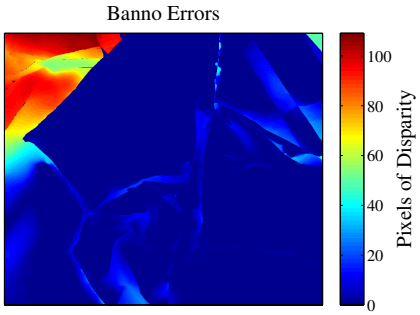
Figure 51: Cloth 2 Confidence Maps. White indicates a high confidence while black indicates a low confidence.



(a) C3D



(b) CIGAD



(c) Banno

Figure 52: Cloth 2 Disparity Errors

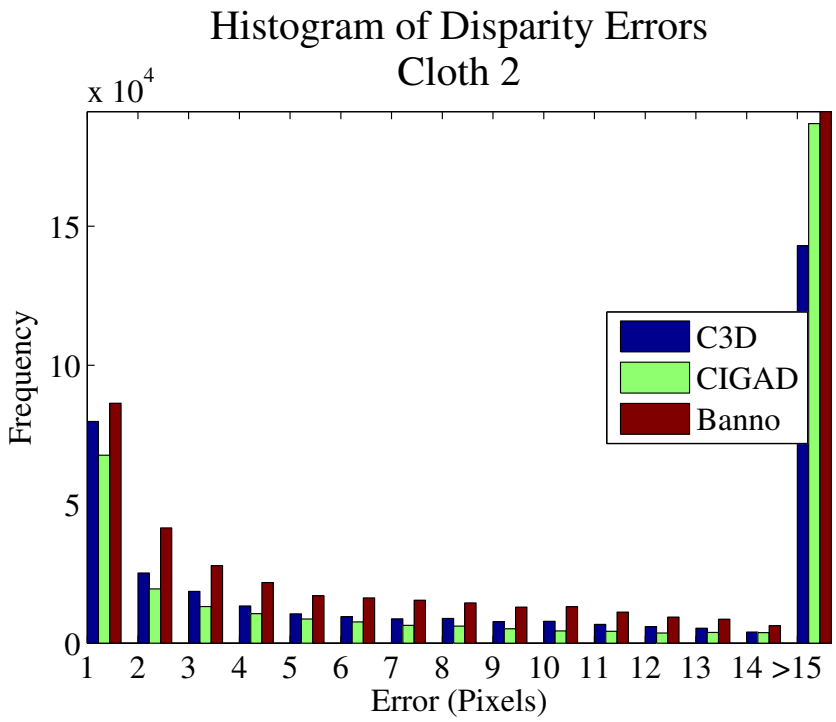
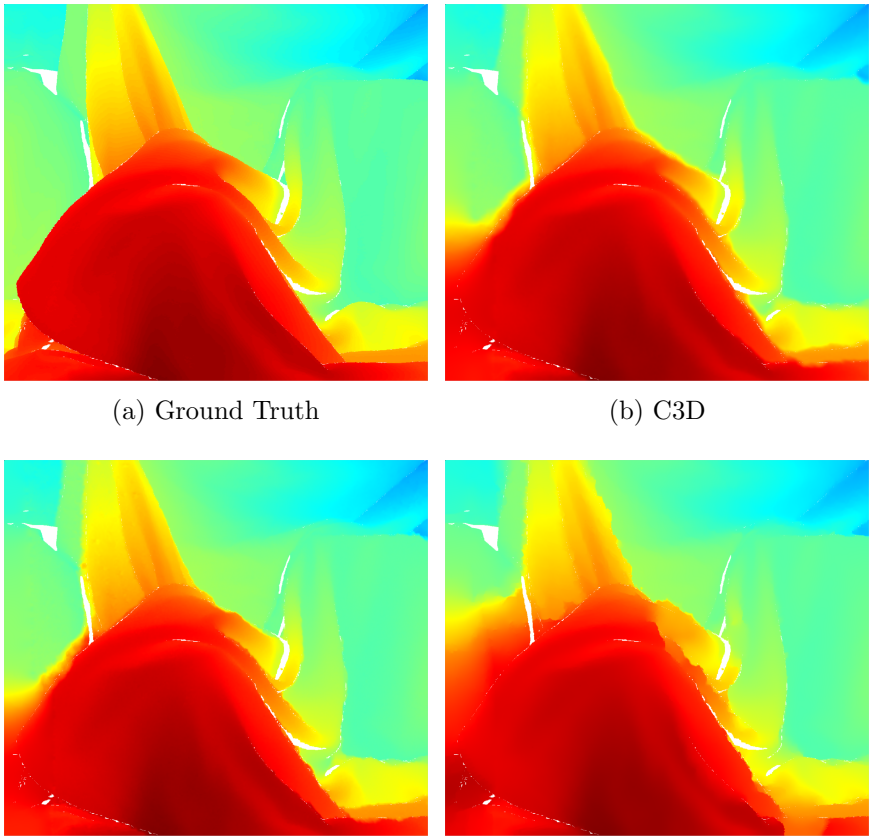


Figure 53: Cloth 2 Error Histogram

A.10 Cloth 3

(a) Left Frame (b) Right Frame

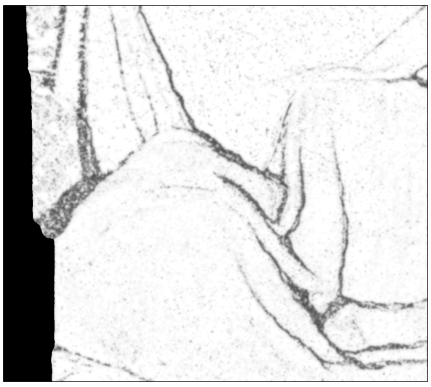
Figure 54: Cloth 3 Input Frames



(a) Ground Truth (b) C3D

(c) CIGAD (d) Banno

Figure 55: Cloth 3 Disparity Maps

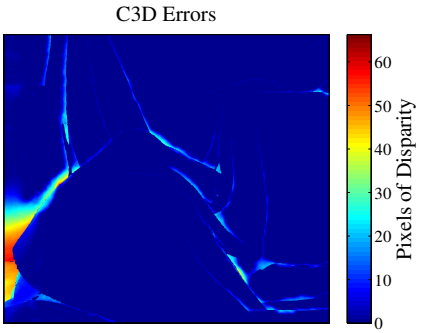


(a) C3D

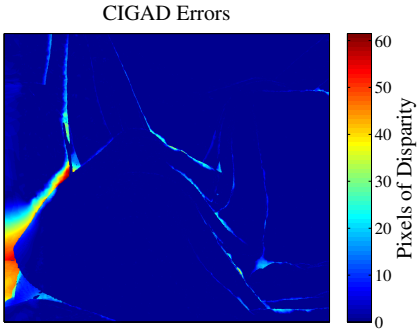
(b) CIGAD

(c) Banno

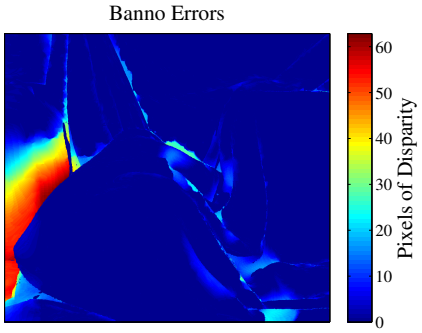
Figure 56: Cloth 3 Confidence Maps. White indicates a high confidence while black indicates a low confidence.



(a) C3D



(b) CIGAD



(c) Banno

Figure 57: Cloth 3 Disparity Errors

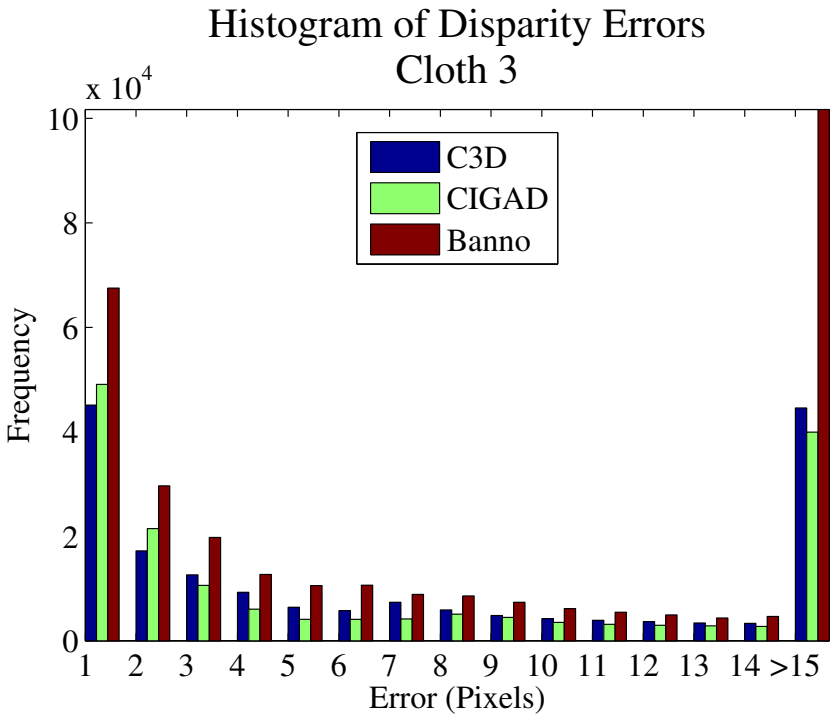
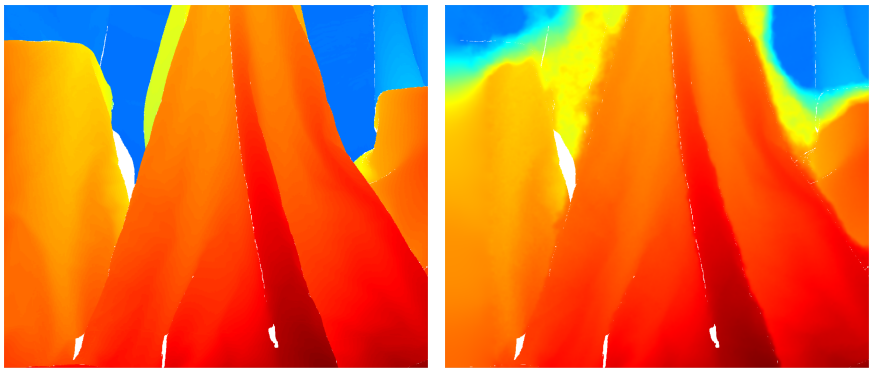


Figure 58: Cloth 3 Error Histogram

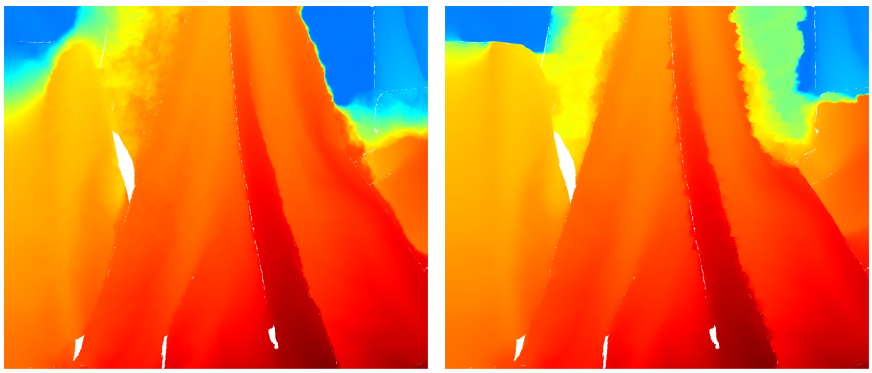
A.11 Cloth 4

(a) Left Frame (b) Right Frame

Figure 59: Cloth 4 Input Frames

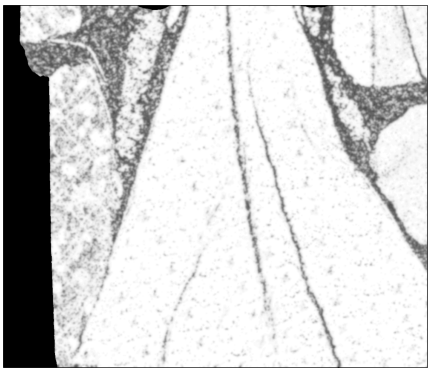


(a) Ground Truth (b) C3D

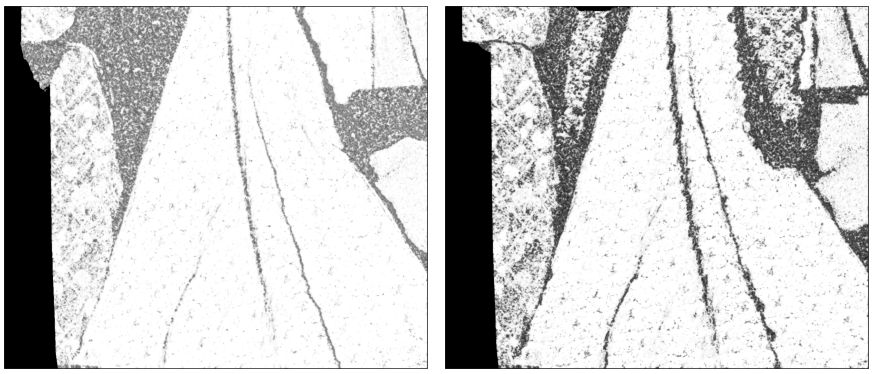


(c) CIGAD (d) Banno

Figure 60: Cloth 4 Disparity Maps

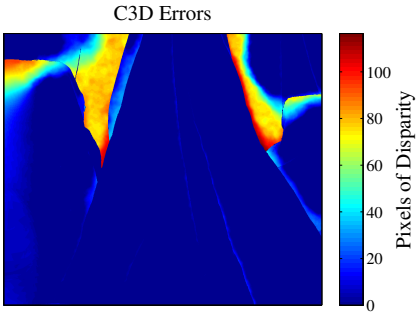


(a) C3D

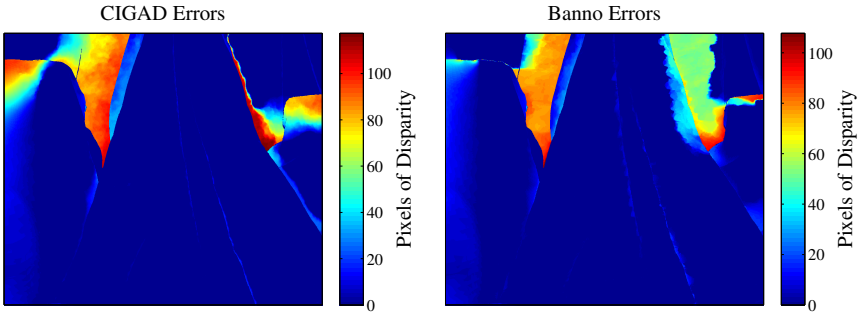


(b) CIGAD (c) Banno

Figure 61: Cloth 4 Confidence Maps. White indicates a high confidence while black indicates a low confidence.



(a) C3D



(b) CIGAD (c) Banno

Figure 62: Cloth 4 Disparity Errors

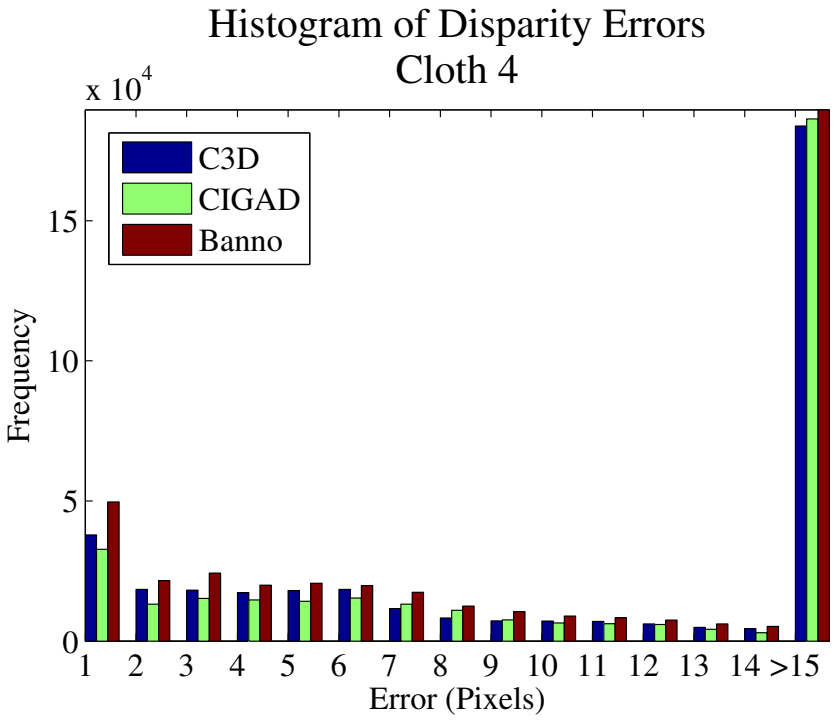


Figure 63: Cloth 4 Error Histogram

Works Cited

[1] S.T. Acton. A pyramidal edge detector based on anisotropic diffusion. In *Acoustics, Speech, and Signal Processing, 1996. ICASSP-96. Conference Proceedings., 1996 IEEE International Conference on*, volume 4, pages 2215 –2218 vol. 4, May 1996.

[2] S.T. Acton. Multigrid anisotropic diffusion. *Image Processing, IEEE Transactions on*, 7(3):280 –291, Mar 1998.

[3] S.T. Acton, A.C. Bovik, and M.M. Crawford. Anisotropic diffusion pyramids for image segmentation. In *Image Processing, 1994. Proceedings. ICIP-94., IEEE International Conference*, volume 3, pages 478 –482 vol.3, Nov 1994.

[4] Jean Babaud, Andrew P. Witkin, Michel Baudin, and Richard O. Duda. Uniqueness of the gaussian kernel for scale-space filtering. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PAMI-8:26–33, 1986.

[5] A. Banno and K. Ikeuchi. Disparity map refinement and 3d surface smoothing via directed anisotropic diffusion. *International Conference on Computer Vision Workshops*, pages 1870–1877, Oct 2009.

[6] S.T. Barnard and M.A. Fischler. Computational stereo. *ACM Computing Surveys*, 14:553–572, Dec 1982.

[7] James R. Bergen, P. Anandan, Th J. Hanna, and Rajesh Hingorani. Hierarchical model-based motion estimation. pages 237–252. Springer-Verlag, 1992.

[8] M.J. Black, G. Sapiro, D.H. Marimont, and D. Heeger. Robust anisotropic diffusion. *Transactions on Image Processing*, 7(3):421–432, Mar 1998.

[9] Y. Boykov, O. Veksler, and R. Zabih. Fast approximate energy minimization via graph cuts. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 23(11):1222–1239, Nov 2001.

[10] M.Z. Brown, D. Burschka, and G.D. Hager. Advances in computational stereo. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 25(8):993 – 1008, Aug 2003.

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

[12] J. Canny. A computational approach to edge detection. *Transactions on Pattern Analysis and Machine Intelligence*, 8:679–698, 1986.

[13] J.J. Clark. Singularity theory and phantom edges in scale space. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 10(5):720–727, Sep 1988.

[14] S Cochran. Toys image, no. 3-2. `http://vasc.ri.cmu.edu//idb/html/stereo/toys/index.html`.

[15] B. Cyganek and J.P. Siebert. *An introduction to 3D computer vision techniques and algorithms*. John Wiley & Sons, Ltd, Jan 2009.

[16] C. Dima and S. Lacroix. Using multiple disparity hypotheses for improved indoor stereo. In *Robotics and Automation, 2002. Proceedings. ICRA '02. IEEE International Conference on*, volume 4, pages 3347 – 3353 vol.4, 2002.

[17] Jingting Ding, Xin Du, Xinhuan Wang, and Jilin Liu. Improved real-time correlation-based fpga stereo vision system. In *Mechatronics and Automation (ICMA), 2010 International Conference on*, pages 104 –108, Aug 2010.

[18] P. Favaro, S. Osher, S. Soatto, and L. Vese. 3d shape from anisotropic diffusion. In *Computer Vision and Pattern Recognition, 2003. Proceedings. 2003 IEEE Computer Society Conference on*, volume 1, pages I–179 – I–186 vol.1, Jun 2003.

[19] P. Fua. A parallel stereo algorithm that produces dense depth maps and preserves image features. *Machine Vision and Applications*, 6(1):35–49, Dec 1993.

[20] A Geiger, M Roser, and R Urtasun. Efficient large-scale stereo matching. *Asian Conference on Computer Vision*, 2010.

[21] M.R. Hajiaboli. An anisotropic fourth-order diffusion filter for image noise removal. *International Journal of Computer Vision*, pages 1–15, 2010.

[22] H. Hirschmuller. Stereo processing by semiglobal matching and mutual information. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 30(2):328 –341, Feb 2008.

[23] X. Hu and P. Mordohai. Evaluation of stereo confidence indoors and outdoors. In *Computer Vision and Pattern Recognition*, pages 1466–1473, Jun 2010.

[24] S. Ince and J. Konrad. Occlusion-aware optical flow estimation. *Image Processing, IEEE Transactions on*, 17(8):1443 –1451, Aug 2008.

[25] J.Z. Jin, T.B. Niblett, and C.W. Urquhart. Improved methods and apparatus for 3-d imaging. Patent Application, May 2000. WO 2000/027131 A2.

[26] T. Kanade and M. Okutomi. A stereo matching algorithm with an adaptive window: theory and experiment. In *Robotics and Automation, 1991. Proceedings., 1991 IEEE International Conference on*, pages 1088 –1095 vol.2, Apr 1991.

[27] Junhwan Kim, V. Kolmogorov, and R. Zabih. Visual correspondence using energy minimization and mutual information. In *Computer Vision, 2003. Proceedings. Ninth IEEE International Conference on*, pages 1033 –1040 vol.2, Oct 2003.

[28] Jan Koenderink. The structure of images. *Biological Cybernetics*, 50:363–370, 1984. 10.1007/BF00336961.

[29] S. Lefebvre, S. Ambellouis, and F. Cabestaing. A colour correlation-based stereo matching using 1d windows. In *Signal-Image Technologies and Internet-Based System, 2007. SITIS '07. Third International IEEE Conference on*, pages 702 –710, dec. 2007.

[30] Maabar M. and Siebert J.P. Smoothing disparity maps using intensity-edge guided anisotropic diffusion. *Medical Image Understanding and Analysis*, Jul 2008.

[31] Y. Nakamura, T. Matsuura, K. Satoh, and Y. Ohta. Occlusion detectable stereo-occlusion patterns in camera matrix. In *Computer Vision and Pattern Recognition, 1996. Proceedings CVPR '96, 1996 IEEE Computer Society Conference on*, pages 371 –378, Jun 1996.

[32] Nalcioglu O. and Cho Z.H. Limits to signal-to-noise improvement by fid averaging in nmr imaging. *Physics in Medicine and Biology*, 29(8):969, 1984.

[33] P. Perona and J. Malik. Scale-space and edge detection using anisotropic diffusion. *Pattern Analysis and Machine Intelligence*, 12(7):629 –639, Jul 1990.

[34] M. Proesmans, L. Van Gool, E. Pauwels, and A. Oosterlinck. Determination of optical flow and its discontinuities using non-linear diffusion. In Jan-Olof Eklundh, editor, *Computer Vision - ECCV '94*, volume 801 of *Lecture Notes in Computer Science*, pages 294–304. Springer Berlin / Heidelberg, 1994. 10.1007/BFb0028362.

[35] Lynn H. Quam. Readings in computer vision: issues, problems, principles, and paradigms. chapter Hierarchical warp stereo, pages 80–86. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 1987.

[36] E. Rifkah and A. Amer. Fast automated stopping-time and edge-strength estimation for anisotropic diffusion. In *Acoustics, Speech and Signal Processing*, pages 933–936, Apr 2008.

[37] H. Scharr, M.J. Black, and H.W. Haussecker. Image statistics and anisotropic diffusion. In *Computer Vision*, volume 2, pages 840–847, Oct 2003.

[38] D. Scharstein and R. Szeliski. High-accuracy stereo depth maps using structured light. In *Computer Vision and Pattern Recognition, 2003. Proceedings. 2003 IEEE Computer Society Conference on*, volume 1, pages I–195 – I–202 vol.1, Jun 2003.

[39] Daniel Scharstein and Richard Szeliski. Stereo matching with nonlinear diffusion. *International Journal of Computer Vision*, 28:155–174, 1998. 10.1023/A:1008015117424.

[40] Daniel Scharstein and Richard Szeliski. A taxonomy and evaluation of dense two-frame stereo correspondence algorithms. *International Journal of Computer Vision*, 47:7–42, Apr 2002.

[41] Victor Solo. A fast automatic stopping criterion for anisotropic diffusion. In *Acoustics, Speech, and Signal Processing*, volume 2, pages 1661–1664, May 2002.

[42] A.K.W. Sum and P.Y.S. Cheung. Stabilized anisotropic diffusions. In *Acoustics, Speech and Signal Processing*, volume 1, pages 709–712, Apr 2007.

[43] R. Szeliski. Stereo correspondence. In *Computer Vision*, Texts in Computer Science, pages 467–503. Springer London, 2011.

[44] R. Szeliski and J. Coughlan. Spline-based image registration. *International Journal of Computer Vision*, 22:199–218, Mar 1997.

[45] Paul Trepagnier, Jorge Nagel, Powell Kinney, Matthew Dooner, Sergey Drakunov, Aaron Lee, Michael Dewenter, Matt Hardey, and Eric Gray. Team gray's 2007 urban challenge vehicle. Technical report, June 2007.

[46] Paul G. Trepagnier, Jorge Nagel, Powell M. Kinney, Cris Koutsougeras, and Matthew Dooner. Kat-5: Robust systems for autonomous vehicle navigation in challenging and unknown terrain. *Journal of Field Robotics*, 23(8):509–526, 2006.

[47] J. Weickert. *Anisotropic diffusion in image processing*. B. G. Teubner, 1998.

[48] A Witkin. Scale-space filtering: A new approach to multi-scale description. *Acoustics, Speech, and Signal Processing, IEEE International Conference on ICASSP '84.*, 9:150–153, 1984.

[49] Y. Yang, A. Yuille, and J. Lu. Local, global, and multilevel stereo matching. In *Computer Vision and Pattern Recognition, 1993. Proceedings CVPR '93., 1993 IEEE Computer Society Conference on*, pages 274 –279, Jun 1993.

[50] Y. You, W. Xu, A. Tannenbaum, and M. Kaveh. Behavioral analysis of anisotropic diffusion in image processing. *Image Processing*, 5(11):1539 –1553, Nov 1996.

[51] Yongjian Yu and S.T. Acton. Speckle reducing anisotropic diffusion. *Image Processing, IEEE Transactions on*, 11(11):1260 – 1270, nov 2002.