Contour Stereo Matching by Correlation in a Constraint Satisfaction Approach

Jochen Lang
Laboratory of Computational Intelligence
Department of Computer Science
University of British Columbia, Vancouver, B.C. V6T 1Z4.
e-mail: jlang@cs.ubc.ca

Abstract

Stereo vision is the recovery of depth of a scene using multiple images with different points of view. The major task in stereo vision is establishing correspondence of points between images given a known pinhole-model of the stereo cameras. Two major approaches to matching points have been used widely: feature-based matching and area- or correlation-based matching. It is well known that both approaches have specific advantages and disadvantages. We describe a method for the matching of contours in a constraint satisfaction approach. Contour matches allow for depth estimation of textured surfaces and objects boundaries to sub-pixel accuracy. These depth estimates may be used for geometric object modeling. We analyze and provide quantative results of the constraint-based stereo matching, and additionally present some examples of objects models obtained with our approach.

1 Introduction

Stereo vision has been extensively studied and several recent overviews exist, e.g. [6, 9] or [16, pp. 217-242] and see also [7] for the numerous related work mentioned. Two approaches to stereo matching are commonly distinguished [9, 16, 20]: intensity-based (sometimes also referred to as area- or correlation-based) matching and feature-based matching. Other approaches to stereo include phase-based matching [10] or energy-based regularization [3, 20].

Feature-based stereo matching establishes correspondence between image features across more than one image. Features which have been used include line segments [2, 14], corners[15] and regions [4]. Individual features need to be compared to select the correct corresponding features. The comparison is based on a feature vector which typically consists of several geometric and photometric properties of the feature. Stereo matching based on intensity[1] establishes correspondence between individual image pixels. Pixels are matched based on the similarity of the brightness pattern of the pixel and its surrounding.

The major advantages of feature-based matching approaches is their potential for high accuracy and reliability at the feature location. However, features are necessarily sparse in an image because they are represented by a distinguished collection of image pixels. Pixels on the other hand are at a fixed density across the entire image.

Contours of objects may be described by features (e.g. line segments) and object contours often indicate a depth discontinuity. Feature-based matching allows for well-localized depth reconstruction at these depth discontinuities. Intensity-based correlation approaches give inaccurate answers exactly at these discontinuities. The inaccuracy is produced by smoothing across the discontinuity due to the size of the correlation window. Also, the correlation function is roughly bimodal [12] with a maximum (or minimum depending on the formulation employed) on either side of the discontinuity.

Unfortunately, line segment approximations to contours are not as distinct as would be desirable and line segment matching is typically plagued by a large number of missing matches, as well as outliers. One of the main sources of error in line segment matching is the approximation of contours by line segments [1]. The apparent contour of an object varies between images of the stereo-view because of the change in view-point. Edge detection by itself is also noisy. In addition, the approxima-

[1] The commonly used term correlation-based stereo is somewhat misleading since feature-based approaches may include a correlation score of the image of the feature as part of their feature vector.

tion of intensity contours by line segments is not stable. Consider a perfectly circular contour to be approximated by line segments; then changing only the initial point for the approximation results in a completely different segmentation.

The constraint satisfaction approach proposed in this paper addresses the errors caused by differing line segment approximations in the different stereo images by segmenting only one of the images. Our constraint-based formulation includes the line-segment approximation only implicitly, therefore avoids this source of error all together. Our approach matches contours approximated by line segments to sub-pixel accuracy. The error in image coordinates is reduced by least-squares line fitting and the disparity is interpolated to achieve sub-pixel accuracy using spline functions.

2 Stereo Matching as Constraint Satisfaction

Constraint satisfaction is a general framework in Artificial Intelligence [13]. A problem is formulated as a set of constraints which have to be met in order for a goal to be achieved. Two type of constraints are commonly distinguished; soft and hard constraint. A hard constraint has to be met for a goal to be established while a soft constraint serves to favor one solution over another.

A solution to the correspondence problem in stereoscopic vision is constrained by the epipolar constraint. For a pinhole camera, a point in an image defines a line in three-dimensions passing through this image point and the nodal point of the camera. However, this imaginary line in three-dimensions maps to a line in the two-dimensional image plane of a pinhole camera at a different position or orientation. This two-dimensional line is the epipolar line of the image point of the first camera in the second camera. The epipolar constraint restricts all matches to be on a line but does not favor any match candidate on that line. The match candidates on this line need to be evaluated with respect to some other constraints. Other possible constraints include the uniqueness constraint, the ordering constraint, various continuity constraints, the disparity gradient constraint, etc. All these constraints are aimed at selecting the proper match given the epipolar constraint. The constraints employed in our approach are discussed in Section 4.

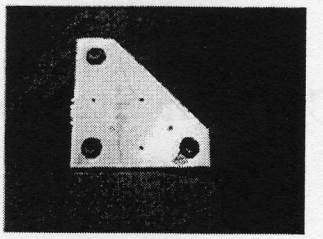


Figure 1: Triclops Stereo-Head

3 The Triclops System

Experiments with our algorithm are conducted with the *Triclops* trinocular stereo-vision system of Point Grey Research [19] (see Figure 1). The vision hardware of the system consists of a module featuring three calibrated cameras. The cameras are arranged in a L-shape configuration focusing at infinity. The baseline of the cameras is about $10cm$ and the focal length is about $3.8mm$. The *Triclops* stereo vision system provides near real-time range images using area-based correspondence. However, for the purpose of this paper only the rectified and calibrated images are used as opposed to the coarse but fairly dense depth maps produced by *Triclops*. The rectification warps the strongly radially-distorted images such that the image-triplet is aligned along the image axis of the reference image. The reference image is the image in the corner of the L-arrangement. The parameters of the pinhole model for the cameras after calibration consists simply of a focal length in image coordinates identical for all three cameras. The two very similar baselines of the horizontal and vertical image pair are also calibrated. The complete calibration is part of the *Triclops* product. The calibration procedure typically results in RMS errors of a fifth of a pixel with a maximum error of half a pixel [19].

4 Algorithm Overview

In this section, we discuss some preliminary steps for the matching algorithm and provide an overview of the algorithm, more details are given in Section 5 and 6.

We assume that the images of a stereo-view are rectified and aligned along the image axis of the reference image. This assumption is commonly made and image warping or rectification can assure that this holds for an L-shape camera configuration [1]. Our experiments are based on the *Triclops* stereo-head including its rectification software, thus the assumption is valid for our hardware.

The first step of our algorithm detects edges of the rectified reference image. The reference image is the image in the corner of the L-arrangement. Intensity edgels are found with the Canny edge detector[5], the edgels are linked into chains which are then approximated by line segments. This processing is accomplished by means of the *Vista* vision software package [17].

The subsequent steps of the algorithm are determined by the constraints employed for the matching. The stereo matcher proceeds in the following steps:

1. Calculate a correlation score for endpoints and midpoint of each line segment in reference image for both, horizontal and vertical image pair.
2. Approximate the continuous correlation function by spline interpolation based on the per-pixel correlation score.
3. Find a unique solution satisfying the line constraint and a soft disparity gradient constraint by means of function minimization.
4. Enforce a global 2-D continuity constraint in the image in order to dismiss outliers.
5. Enforce a global 3-D continuity constraint for connected edge contours to improve accuracy.

The following Sections discuss the details of the above steps in terms of the relevant constraints.

5 Local Constraints

5.1 Line Constraint

The major advantage of our algorithm compared to other feature-based matching algorithms is the use of a line constraint instead of an explicit feature representation. The line constraint restricts matches along a line segment in one image to a line in another image of a set of stereo-views. This is very different from matching line segments which approximate an intensity contour in one image to another set of line segments which in turn approximate an intensity contour in another image. In general, the line constraint holds if a 3-D line segment has given rise to the 2-D line segment in the image. The constraint also approximately holds if the viewpoint of the stereo-views is not too different or the 3-D curve is well approximated by a 3-D line segment. The assumption of a 2-D line segment in an image being caused by an (approximate) 3-D line segment is common to all line segment based matching algorithms. See Section 5.3 for the implementation of the line constraint.

5.2 Further Local Constraints

The photometric constraint as employed in our algorithm states that a pixel in one image has a surrounding window with a similar brightness pattern in another image of the stereo-view. A window in one image of the stereo pair is compared to a window in the other image at various disparities. The constraint is evaluated by the window-normalized sum of squared differences as in [8] which is zero for identical windows in the stereo-pairs.

Furthermore, the algorithm employs a strong uniqueness constraint. The uniqueness constraint states that one feature can only be at a unique disparity. In other words, a line segment can have only one true matching line segment in another image of the stereo-view. The uniqueness constraint is false if features line-up in one image.

The algorithm also employs a soft disparity gradient constraint. The soft disparity gradient constraint favors lower disparity gradients over larger ones. Some justification of the constraint has previously derived from psychological research [1]. However, the main justification of the constraint derives from the improved robustness of the matching. A large disparity gradient of a line segment indicates a large slopes in the direction of the line-of-sight. The localization of such a line segment is necessarily very poor. The extreme case of the segment being collinear with the line-of-sight illustrates this drastically. All points on such a line map into the same image pixel and no constraint on the length of the line or its orientation can be derived from this image.

5.3 Formulation of Minimization

The above discussed local constraints are evaluated by solving a continuous optimization problem based on correlation scores. The correlation scores are calculated along the epipolar lines over the valid disparity range at the endpoints and the midpoint of the line segment of the reference image. The line constraint requires the midpoint to be collinear with the endpoints of the segment. The soft photometric constraint requires the sum of the correlation scores to be as close to zero as possible. The hard uniqueness constraint allows only a single *minimal* solution. The soft disparity gradient will favor solutions for which the disparity gradient of the line is low. A solution may be found by minimizing the error term of Equation 1.

$$Error = \sum_{pt=1}^{3} \sum_{pair=1}^{2} \left(w_{(pair,pt)} * f^2_{(pair,pt)}(d_{pt}) \right) + w_{gradient} * (d_1 - d_2)^2 \quad (1)$$

d_1:	Disparity at endpoint 1 of the line segment
d_2:	Disparity at endpoint 2 of the line segment
d_3:	$0.5\ (d_1 + d_2)$
$pair_1$:	Horizontal pair of stereo-images
$pair_2$:	Vertical pair of stereo-images
$f_{(pair,pt)}$:	Correlation score at point pt in a given pair
$w_{(pair,pt)}$:	Weight of correlation score
$w_{gradient}$:	Weight of gradient constraint
$Error_{max}$:	1.0

Three major tasks remain; finding a continuous approximation to the correlation score *sampled* at pixels (see Section 5.3.1), assigning weights to the components of Equation 1 (see Section 5.4) and minimizing the equation.

The minimization can be numerically solved for disparities d_1 and d_2 by the conjugate gradient method [18, pp. 420-425]. The method requires the partial derivatives of Equation 1 and a *good* guess of the solution. The solution may be guessed as the minimum of the discrete correlation scores of only the endpoints of the line-segment in the reference frame. For the trinocular case, each the horizontal and vertical image pair contribute a *possibly* different guess for disparities d_1 and d_2. The minimization proceeds from the initial guess with the lowest error. We only accept solution with an error of less than one.

5.3.1 Spline Interpolation

Spline interpolation approximates a sampled curve by piece-wise low-order polynomials. We use cubic splines, i.e. each polynomial is of third order. A cubic is uniquely defined by four samples, in other words between any two consecutive samples are two degrees of freedom left for the spline. Natural cubic splines require that the first and second derivatives are identical at the sample points for both cubics attached to any specific sample point. This requirement leads to a tri-diagonal linear system of the size of the number of samples. But even more devastating, cubic splines lead to a *very* continuous interpolant with severe ringing at large changes of the second derivative. This ringing has to be damped for meaningful interpolation of the correlation score. We limit the ringing by employing a minimum filter (since we are only interested in the curve around its known maximum of around zero) and the use of cubic tension spline [11] instead of a natural spline.

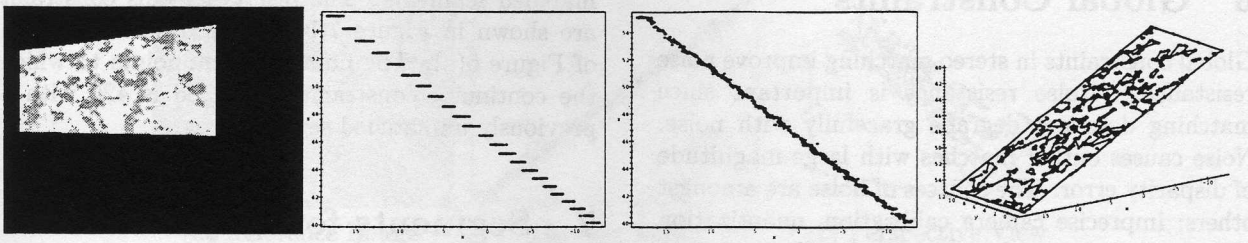
The tension parameter of the tension spline controls the ringing. A very high tension leads to a piecewise linear interpolation between sample points, a tension of 0 leads to a Catmull-Rom spline and negative tension values cause the spline to approximate a cubic spline. The spline is defined completely locally, i.e. each interpolant cubic only depends on the four surrounding samples. We found that a good choice for our task is to set the parameter to 0.4.

5.4 Experimental Results

The experiments described in this section are conducted with a tilted plane with a random pattern and with a checker board pattern. The random pattern is shown in Figure 2(a) as a rendered view and in Figure 6(a) as seen from the reference camera of the *Triclops* stereo-head. The checker board pattern is shown in Figure 5(a). The random pattern is chosen because of its lack of line structure, and the checker board because all of its contours are line segments. Notice that the matcher performs equally well on both type of patterns. We calculate below the RMS error (root of the sum of squared differences from the mean) of the endpoints of the line segments from the plane of the pattern. The RMS serves our comparison but is not very meaningful as an absolute value. The accuracy of the stereo triangulation depends quadratically on the distance. (The standard deviation for the Canny detector is set in all experiments throughout this paper to $\sigma = 0.5$.)

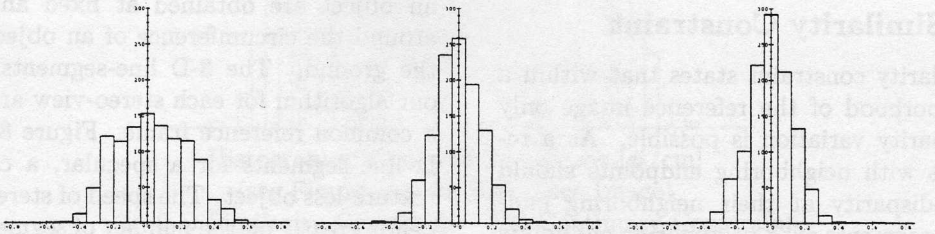
Figure 2 shows the effect of the spline interpolation on the depth value. The pattern is at fixed intervals without interpolation. The natural spline and the tension spline interpolation perform similar. (For the histograms of the endpoint deviation see Figure 3 and Table 1).

Figure 4 illustrates the effect of the local constraints on a *Triclops* stereo-view (see Figure 6(a)). The *ground-truth* for the plane of the pattern is unknown and therefore, is first estimated by a linear least square procedure. The results show clearly the superiority of the trinocular versus the binocular algorithm. All results at this stage are contaminated by outliers; however the global constraints will address that problem (see Section 6). For the detailed results, see Figure 4 and Table 1. In the case of the *Triclops* stereo-view (Figure 6(a)) the RMS is the lowest if the line constraint is not employed. The line constraint is violated by the random pattern. The line constraint performs well on the checker board (see Figure 5 and Table 1).



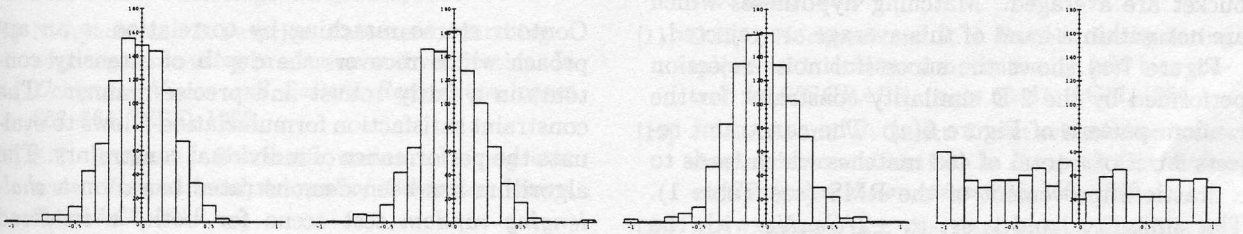
From Left to Right: (a) Reference Image (b) Top View No Interpolation
(c) Top View Tension Spline (d) Perspective View Tension Spline

Figure 2: Tension Spline Interpolation with Rendered Stereo-View



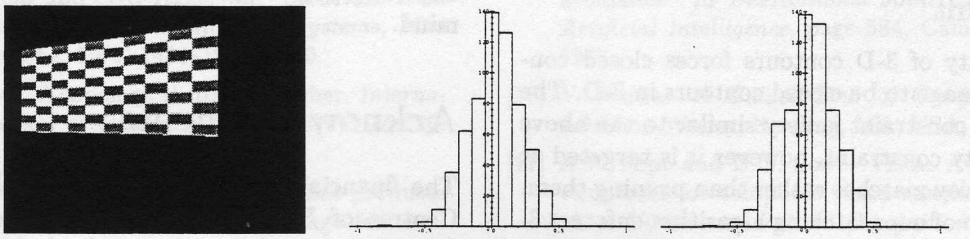
Histogram of Endpoint Deviations [#/cm]
(a) No Interpolation (b) Natural Spline (c) Tension Spline
(see Figure 2(a) for Reference Image)

Figure 3: Choice of Spline Interpolation



Histogram of Endpoint Deviations [#/cm]
(a) All Local Constraints (b) No Disp. Grad. Constraint (c) No Line Constraint (d) Binocular View
(see Figure 6(a) for Reference Image)

Figure 4: Effect of Local Constraints on *Triclops* Stereo-View



Histogram of Endpoint Deviations [#/cm]
(a) Reference Image (b) No Line Constraint (c) With Line Constraint

Figure 5: Effect of Local Constraints on Rendered View

6 Global Constraints

Global constraints in stereo matching improve noise resistance. Noise resistance is important since matching does not degrade gracefully with noise. Noise causes outlier matches with large magnitude of disparity error. The sources of noise are amongst others; imprecise camera calibration, quantization of image location and intensity, as well as geometric and photometric effects not accounted for by the matching strategy. For example, our stereo matcher does not explicitly handle occlusions, shadows or highlights.

6.1 2-D Similarity Constraint

The 2-D similarity constraint states that within a certain neighborhood of the reference image only a certain disparity variation is possible. As a result, segments with neighboring endpoints should have similar disparity at their neighboring endpoints. These segments with neighboring endpoints are found with a bucketing technique applied to the segments of the reference image. The implementation assumes that within a bucket only one disparity is valid for endpoints. Note, that these endpoints commonly belong to the same intensity contour. Therefore, all the disparity hypothesis in the bucket are averaged. Matching hypotheses which are not within a band of this average are rejected.

Figure 7(a) shows the successful noise rejection performed by the 2-D similarity constraint for the random pattern of Figure 6(a). The constraint rejects 9 out of a total of 468 matches which leads to a drastic improvement of the RMS (see Table 1). The similarity band is set to $\pm 2/SegNo$ with the number of other segments $SegNo$ in the bucket of size 4. The constraint is very successful in rejecting the outliers and preserving the other matches. The density of the pattern aids the constraint.

6.2 3-D Contour Continuity Constraint

The continuity of 3-D contours forces closed contours in an image to be closed contours in 3-D. The effect of this constraint is very similar to the above 2-D continuity constraint, however it is targeted at improving noisy matches rather than pruning them. This last step of our matching algorithm enforces 3-D continuity by following edge chains. This step has two effects: additional matches are found for previously unmatched segments with a matched predecessor and successor in the chain; and second, the disparity is averaged for two joining previously matched segments. The results for this constraint are shown in Figure 7(b) for the random pattern of Figure 6(a). The number of endpoints to which the continuity constraint is applied is 420 with 4 previously unmatched segments.

7 Segments for Object Modeling

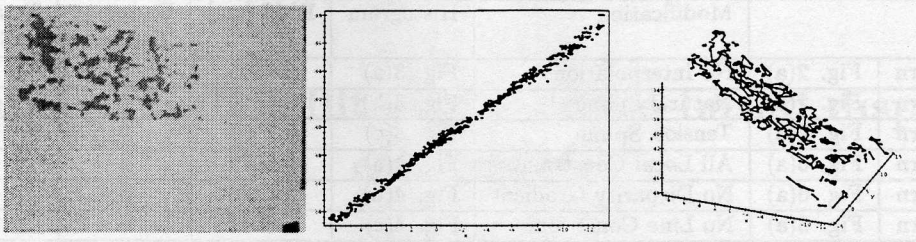
In this section, we briefly demonstrate the use of our algorithm for object modeling in a shape from rotation approach [21]. Registered stereo-views of an object are obtained at fixed angular spacing around the circumference of an object coplanar to the ground. The 3-D line-segments produced by our algorithm for each stereo-view are rotated into a common reference frame. Figure 8 shows the 3-D line segments for a specular, a concave and a texture-less object. The speed of stereo matcher depends greatly on the number of segments in the reference image. Typical values on our network server (*Sun Sparc Ultra-2*) are in the order of 2 to 8 stereo-triplets per minute.

8 Conclusion

Contour stereo matching by correlation is an approach which recovers the depth of intensity contours in a fairly robust and precise manner. The constraint satisfaction formulaziation allows to evaluate the performance of individual constraints. The algorithm has been demonstrated based on a *challenging random* test scene for both, a rendered stereo-view and a *Triclops* stereo-view. Researchers interested in employing *Triclops* for their own investigations may find the error levels achieved informative. Additionally, we have provided some results for *challenging* objects from several views. These results may serve as a starting point for obtaining object models which is the application we have in mind.

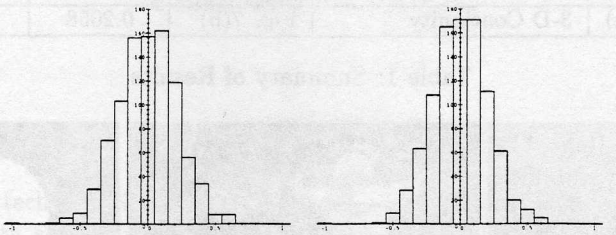
Acknowledgments

The financial support of the Federal Networks of Centres of Excellence IRIS project and NSERC Canada is gratefully acknowledged. My thanks go to Dr. R.J. Woodham for his supervision of this work; D. Murray for many discussions, and R. Barman, C. Jennings, S. Kingdon, D. Murray, V. Tucakov from PGR for their technical support.



(a) Reference Image (b) Top View (c) Perspective View

Figure 6: Effect of Global Constraints



Histogram of Endpoint Deviations [#/cm]
(see Figure 6(a) for Reference Image)
(a) 2-D Similarity Constraint (b) Continuity Constraint

Figure 7: Global Constraints

References

[1] N. Ayache. *Artificial Vision for Mobile Robots.* MIT Press, Cambridge, MA, USA, 1991.

[2] N. Ayache and B. Faverjon. Fast stereo matching of edge segments using prediction and verification. In *Computer Vision and Pattern Recognition*, pages 662–664. IEEE, 1985.

[3] S.T. Barnard. Stochastic stereo matching over scale. *International Journal of Computer Vision*, 3(1):17–32, 1989.

[4] K.L. Boyer and A.C. Kak. Structural stereopsis for 3-d vision. *Transaction on Pattern Recognition and Machine Intelligence*, 10(2):144–166, 1988.

[5] J. Canny. A computational approach to edge detection. *Transaction on Pattern Recognition and Machine Intelligence*, 8(6):679–698, 1986.

[6] U.R. Dhond and J.K. Aggarwal. Structure from stereo - a review. *Transactions on Systems, Man and Cybernetic*, 19(6):1489–1510, 1989.

[7] M.M. Fleck. A topological stereo matcher. *International Journal of Computer Vision*, 6(3):197–226, 1991.

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

[9] M.R.M. Jenkin, A.D. Jepson, and J.K. Tsotsos. Techniques for disparity measurement. *Computer Vision, Graphics and Image Processing: Image Understanding*, 53(1):14–30, 1991.

[10] A.D. Jepson and M.R.M. Jenkin. The fast computation of disparity from phase difference. In *Computer Vision and Pattern Recognition*, pages 398–403. IEEE, 1989.

[11] D.H.U. Kochanek and R.H. Bartels. Interpolating splines with local tension, continuity, and bias control. In *SIGGRAPH*, pages 33–41. ACM, 1984.

[12] J.J. Little and W. Gillett. Direct evidence for occlusion in stereo and motion. In *European Conference on Computer Vision*, pages 336–340, 1990.

[13] A.K. Mackworth. *The Encyclopidia of AI*, chapter Constraint satisfaction. John Wiley, New York, USA, 2 edition, 1991.

[14] G. Medioni and R. Nevatia. Segment-based stereo matching. *Computer Vision, Graphics and Image Processing*, 31:2–18, 1985.

[15] H.P. Moravec. Towards automatic visual obstacle avoidance. In *International Joint Conference on Artificial Intelligence*, page 584, Cambridge, MA, 1977.

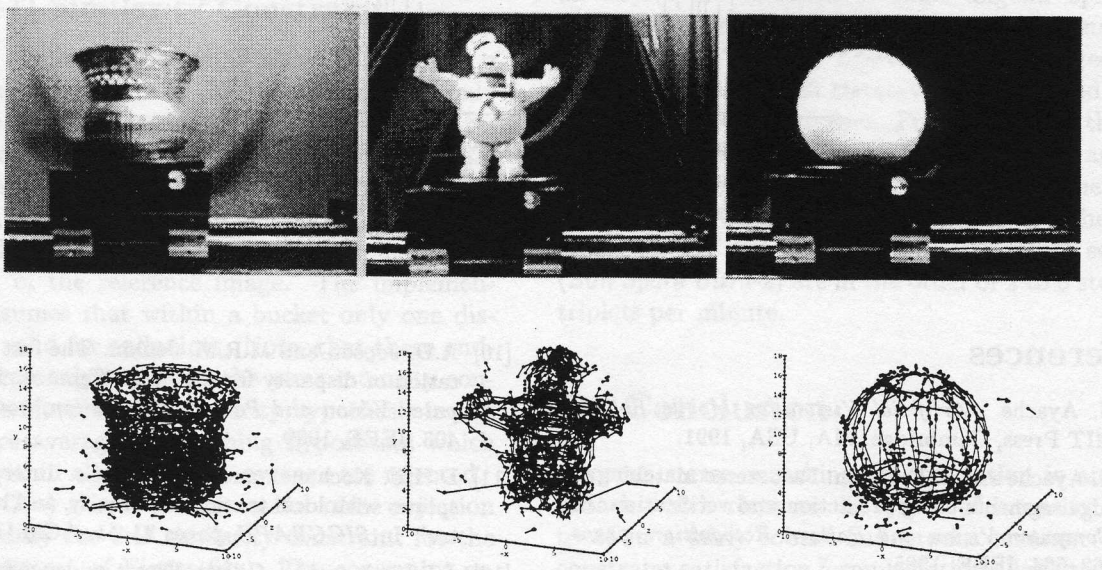
[16] V. Nalwa. *A Guided Tour of Computer Vision.* Addision-Wesley, Reading, MA, USA, 1993.

[17] A.R. Pope and D.G. Lowe. Vista: A software environment for computer vision research. In *Computer Vision and Pattern Recognition*, pages 768–772. IEEE, 1994.

[18] W.H. Press, S.A. Teukolsky, W.T. Vetterling, and B.P. Flannery. *Numerical recipes in C: the art of scientific computing.* Cambridge University Press, 2 edition, 1992.

Scene		Modification	Histogram	RMS [cm]	Endpoint Outliers	Total Matched Segments
Random pattern	Fig. 2(a)	No Interpolation	Fig. 3(a)	0.2103	0	551
Random pattern	Fig. 2(a)	Natural Spline	Fig. 3(b)	0.1410	0	551
Random pattern	Fig. 2(a)	Tension Spline	Fig. 3(c)	0.2103	0	551
Random pattern	Fig. 6(a)	All Local Constraints	Fig. 4(a)	0.6979	4	472
Random pattern	Fig. 6(a)	No Disparity Gradient	Fig. 4(b)	0.7226	8	472
Random pattern	Fig. 6(a)	No Line Constraint	Fig. 4(c)	0.6147	6	472
Random pattern	Fig. 6(a)	Binocular View	Fig. 4(d)	2.4191	671	714
Square pattern	Fig. 5(a)	All Local Constraints	Fig. 5(b)	0.1918	0	309
Square pattern	Fig. 5(a)	No Line Constraint	Fig. 5(c)	0.3715	1	309
Random pattern	Fig. 6(a)	2-D Similarity	Fig. 7(a)	0.2595	0	459
Random pattern	Fig. 6(a)	3-D Continuity	Fig. 7(b)	0.2058	0	463

Table 1: Summary of Results



(a) Brass Vase (b) Stay-Puft (c) Sphere

Top Row: Reference Image, Bottom Row: 3-D Segments from 12 Views

Figure 8: Segments for Various Objects

[19] Point Grey Research. Triclops on-line manual. http://www.ptgrey.com/.

[20] L. Robert and R. Deriche. Dense depth map reconstruction: A minimization and regularization approach which preserves discontinuities. In *European Conference on Computer Vision*, volume I, pages 439–451, 1996.

[21] R. Szeliski. Shape from rotation. Technical Report 90/13, Digital Equipment Corporation, Cambridge Research Lab, Dec 1990.