Tina Memo No. 2001-012
Presented at BMVC 2001

TINA 2001: The Closed Loop 3D Model Matcher

A. J. Lacey, N. A. Thacker, P. Courtney and S. B. Pollard

Last updated
20 / 02 / 2002

WWW.TINA-VISION.NET

Imaging Science and Biomedical Engineering,
School of Cancer and Imaging Sciences,
University of Manchester, Stopford Building,
Oxford Road, Manchester M13 9PT, U.K.

TINA 2001: The Closed Loop 3D Model Matcher

A. J. Lacey, N. A. Thacker, P. Courtney and S. B. Pollard[1]
Imaging Science and Biomedical Engineering,
School of Cancer and Imaging Sciences,
University of Manchester, Stopford Building,
Oxford Road, Manchester M13 9PT, U.K.
[1]Stephen Pollard is with Hewlett Packard Labs, Bristol
neil.thacker@manchester.ac.uk

Abstract

This paper discusses advances made to the 3D geometrical model matching system within the TINA machine vision environment over the last 10 years including the recent inclusion of a verification stage. We show how this final step closes the loop on the object location system allowing the theoretical location performance to be attained, eliminating key assumptions used in the existing forward pass algorithms. We explain how the use of a system approach has been crucial to the development of key components and summarise our findings.

1 Introduction

In the late 1980's researchers from the Artificial Intelligence Vision Research Unit (AIVRU) at the University of Sheffield began developing the TINA vision system. The initial focus of this endeavour was the development of a 3D wireframe model matching system (3DMM) which could locate a known object in a scene from stereo camera data, accurately enough to guide a robotic arm. Tackling such a task required solutions to several key problems in machine vision including, feature extraction and fitting, depth from stereo, automatic stereo camera calibration and robust view based model matching. The algorithms developed were published independently as solutions to the problems they targeted, but all were developed within the unifying framework of TINA. Development of the entire TINA vision system continued to extend existing solutions as well as research techniques to tackle new image understanding problems. In all this time we have endeavoured to ensure that the TINA infrastructure itself remains an environment suitable for sustainable vision research.

Recently we have returned to the 3D wireframe model matcher with a view to verifying the object location and assumed camera model. Model verification has been used in the TINA model matching system before, as a means of rejecting hypothesis [15]. However, we have been able to formulate a solution which closes the loop on the entire existing processing architecture, converting the forward pass algorithm into a fully closed loop iterative solution, eliminating many of the assumptions needed in the original system.

2 The Original 3D Model Matching System

The original version of the 3D model matcher was presented in [12] and [9]. Briefly the system used a sparse edge based depth map extracted from pairs of binocular stereo images together with the corresponding camera calibration information. A geometric interpretation of the scene was constructed by fitting lines and arcs to the depth map data. Statistical matching of 3D scene descriptions to a stored wireframe model enabled the location of the model within the scene to be identified. The algorithm is summarised in the diagram of Fig. 1 and Table 1.

The algorithm of PMF [7] was used to generate the edge based depth maps from the stereo data. Canny edge information [2] was extracted from both images with sub-pixel accuracy (0.1 pixels by experiment). The two extracted edge maps were rectified using the recovered camera calibration to transform the data into a parallel camera system enabling the epi-polar constraint to be exploited. The PMF algorithm solves the stereo correspondence problem using this constraint to closely approximate edge raster correspondence. Disambiguation was further enhanced using edge contrast and orientation consistency as well as local support in the form of disparity gradient information. The disparity gradient limit, which enforces locally similar disparities between neighbouring features, has been shown to have a strong psychophysical basis [1, 10]. The algorithm was formulated to work on

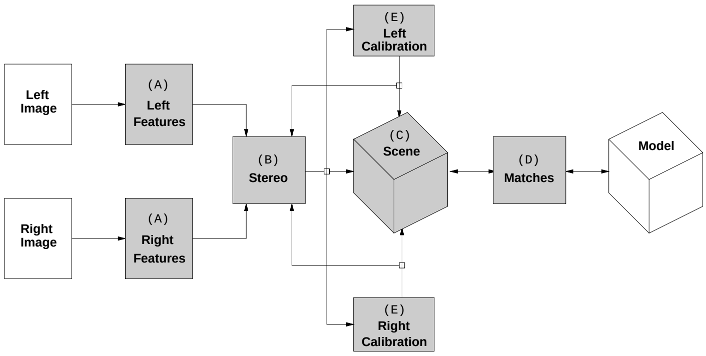


Figure 1: Block diagram of the original 3D model matcher system.

Key	Purpose	Algorithms	Assumptions
A	Edge & Corner Detection 2D Geometry Fitting	Canny [2]	Edges present in expected locations Curves and lines can be correctly linked and fitted
B	Stereo Matching 3D Geometry Fitting	PMF GDB	Accurate epi-polar geometry and match metrics Accurate camera calibration
C	Sequential Model Building	GEOMstat	Accurate feature locations
D	Wireframe Model Matcher	SMM	Gaussian errors on all extracted features Closed form solution is appropriate
E	Camera Calibration	Tsai [4]	Known calibration object present

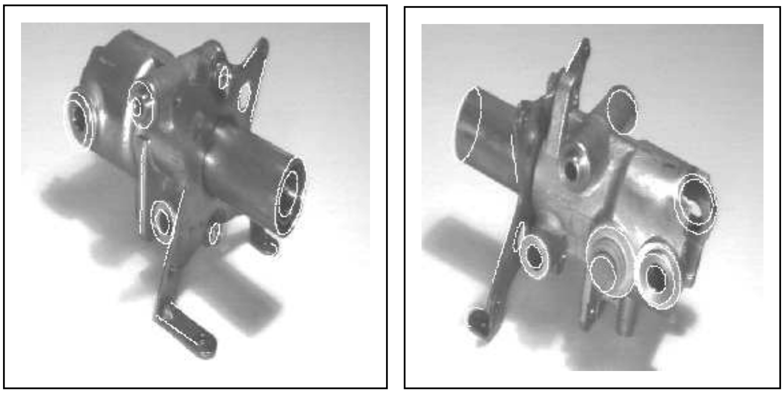
Table 1: Algorithmic descriptions and assumptions for the original 3DMM with key for shaded components in Fig. 1.

the high-level image feature data, i.e. abstract representations of edge features. Camera calibration was achieved using the Tsai algorithm [4].

CONNECT [14] pre-processed the edge based depth map to produce a list of connected edge elements. This list was recursively processed by the GDB algorithm (Geometric Descriptive Base), which aims to classify the list of elements as either a straight line, plane or surface curve. The algorithm attempts first, to fit a straight line to the edge data using orthogonal regression. If this hypothesis is proved unsatisfactory the data is then tested for planarity. If proven, the algorithm attempts to fit a circle to the projection of the data on that plane. If this fails the data is recursively segmented. The resulting geometric interpretation of the data was thus not the only possible representation, but was assumed at least to be reproducible and was adapted to maximise the subsequent use of the data for matching. This was achieved in the Scene Model Matcher, which made some sub-optimal assumption regarding the errors on the geometry in order to generate a rapid closed form solution. It was intended that the geometric data from a scene would be combined over time within a reasoning framework referred to as the REVgraph. This would enable geometric reasoning, sufficient for model building, to be handled by a statistical combination package known as GEOMstat.

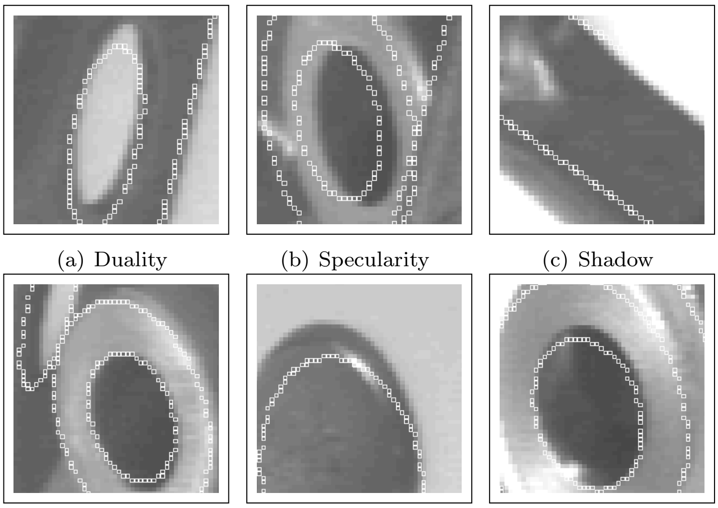
3 Experiments I

Figures 2(a) and 2(b) show two views of an industrial component on which the wireframe model, located with the original 3DMM has been overlaid. Although the location of the model appears reasonable the images of Figs. 3(a)-3(f) demonstrate the inaccuracy of the final model location. The overall inaccuracy in the location arises from many sources. The detail in Fig. 3(a) appears to present a single, elliptical feature. In fact plate thickness gives rise to two elliptical features but only the outer is contained in the model. Details 3(b), 3(d) and 3(f) demonstrate examples of specularity which destroys feature integrity. In detail 3(c) illumination has presented a shadow as a false edge and in 3(e) the elliptical feature is self-occluding. It must be noted that the specific mis-alignments shown in each of the details do not arise directly from the error associated with that region. Rather, those errors result in a compound failure of the model alignment which can generally be attributed to the lack of conformity



(a) Object and located model overlaid - front view (b) Object and located model overlaid - rear view

Figure 2: Typical performance of the original 3DMM on an industrial component.



(a) Duality (b) Specularity (c) Shadow

(d) Specularity (e) Self-Occlusion (f) Specularity

Figure 3: Magnified details from the images in Fig. 2.

of the data to the assumptions listed in Table 1.

4 The Updated 3D Model Matching System

Overall the original 3D model matcher was able to locate a model in the scene to an accuracy of around $5mm$. This was significantly less than the sub-millimetre accuracy which was intended (see Fig. 3). The system was also not as reliable as wished, lacking robustness under illumination changes representative of real world conditions. Finally, some of the algorithmic components, in particular PMF stereo, were formulated in a form which inhibited performance and extendibility. These issues have been addressed by the modifications described in this section and the verification stage introduced in Section 5.

4.1 Stretch correlation

A major development of the new 3DMM system is in the stereo processing algorithms. The stretch correlation algorithm [5] is an area based solution to the stereo problem, matching discrete blocks in the left image to blocks in the right. However, focusing the matching process on information rich areas of the image (those containing edges) improves the robustness of the approach. Most area based correlation techniques assume front-o-parallel surfaces, an assumption which compromises the correlation process by failing to account for surface rotations caused by the changing views. The stretch correlation technique models these rotations by warping, stretching or shearing the right image blocks, which improves the stability of the correlation process. This in turn improves the accuracy of the resulting disparity estimates.

The above technique results in pixel accurate correspondences. Projecting back to the coincident Canny detected edge enables the sub-pixel location of the disparity to be computed, and so the correlation approach is just as accurate as PMF. However, it is more amenable to hardware implementation than PMF, due to the greater homogeneity of the processing. Indeed we have built a prototype VLSI device [6] capable of executing a version of the algorithm at several frames a second. This device is also capable of performing near frame rate image rectification. One final benefit of the stretch correlator is the ease with which it extends to temporal stereo images. We have shown that output from the stretch correlator can be used to bootstrap the search process within a temporal loop [3]. Not only does this reduce the computational requirements of the algorithm but also improves the robustness as the probability of mismatch reduces in proportion to the search area. The REV graph system, which although developed for the original TINA 3DMM was never integrated, was designed to enable fusion of data from temporal-stereo sequences. This system has been made redundant by the temporal aspects of the stretch correlation algorithm.

4.2 Geometry extraction

Several limitations in the geometric processing of the original 3DMM were identified:

- experiments with CONNECT demonstrated unreliability in the context of a working system;
- the Scene Model Matcher (SMM) was sub-optimal and prone to failure. This was mainly attributed to the inability of CONNECT to correctly segment features and the lack of use of appropriate (or even robust) statistics in the optimisation scheme;
- GEOMSTAT didn't cope well with typical variations in image formation (specularity, differences in camera specification and performance) leading to inaccurate ($5mm$ errors) object location;
- calibration of the camera system from a tile was impractical given that the system required continual re-calibration due to drift and physical movement.

As a consequence of these conclusions several efforts were made to incrementally improve performance. Curves and point features were added to the geometrical primitives. It was soon established that a feature based (lines and ellipses) approach [11] was a more reliable method of achieving 3D representations of objects than connectivity based approaches. The edge feature detection process was therefore rationalised and simplified and a 2D version of CONNECT implemented [8] and disparity data fitted to the 2D features.

4.3 Camera calibration

Camera calibration is now achieved within a unified framework which makes use of the data available in a working system. This framework operates in the context of an iterative robust optimisation with optimal combination of covariances via the use of a regularisation term. It is able to track an existing calibration integrating statistical information from known epi-polar errors, robot motion and calibration targets [17]. It was also extended to support calibration of a 4DOF stereo head system [16].

5 Closed-loop Validation

Both the original and updated 3D model matching systems can be regarded as hypothesis generators for object location. Under the assumption of the presence of a particular object in the scene, they return the most probable location and orientation of the object given the evidence. The 3D model matcher uses only the 3D extracted features to achieve this. This leads to error due to a) inaccuracies or failings in segmentation which generate incorrect feature geometry, b) incorrect feature matches and c) simplistic treatment of feature primitive statistics.

Algorithmic improvements in the updated scheme aim to improve the robustness of this system however, a similar level of inaccuracy remains. The ultimate location accuracy of the working system was never consistent with that expected from sub-pixel location of multiple edge features.

The new Closed Loop Validation stage (CLV) closes the loop on the process, testing the generated hypothesis against the original image data and refining this estimate without the constraints imposed by the previous algorithmic stages. An object verification stage [15] was introduced previously in order to asses the quantity of 3D

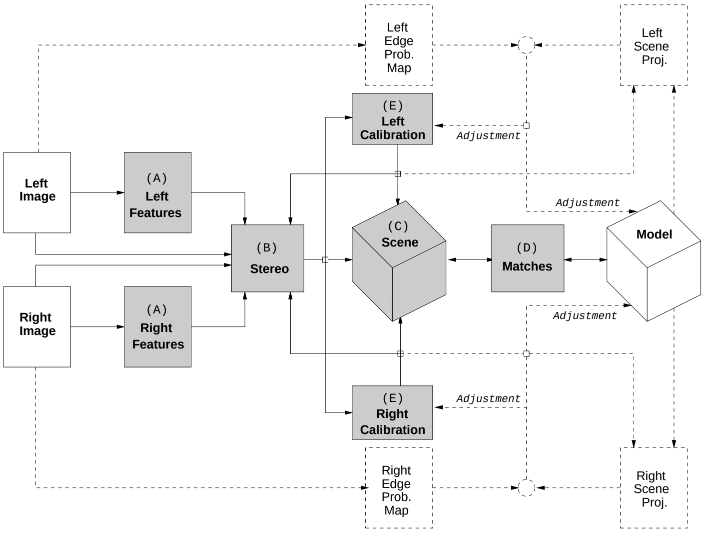


Figure 4: Block diagram of the updated 3D model matcher. The dashed sections represent the additional processing of the closed-loop validation. The letters reference Table 2

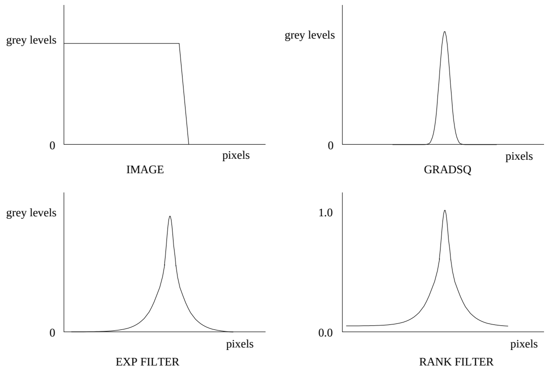


Figure 5: Robust edge probability estimation using image filters. Edges are enhanced using a squared-gradient filter. The exponent and ranking improve robustness and produce relative probabilities.

matches as well as the geometric consistency of the object hypothesis. However, the goal of this system was solely to reject a potentially invalid model hypothesis, providing feedback only within the 3DMM stage. In contrast the CLV compares the final model location with the original image data and is therefore able to feedback over all stages of the algorithm, removing the assumptions used to generate the initial hypothesis. Figure 4 outlines how the CLV stages (dashed lines) compliment the original model matcher (solid lines).

In order to validate the model hypothesis the stereo images are processed to produce an edge probability map where each pixel corresponds to the probability that a world edge was projected onto it. This is achieved by using a combination of filters which generate values which have a maximum value of 1 at an edge and an exponential decay over a scale of several pixels. This approximation is important in order to give the location process the statistical validity required to estimate covariances, Fig. 5.

Given the orientation and location of the model postulated by the original, forward pass of the matcher it is possible to generate a likelihood score for the projection of the located model features given the estimated camera geometry. An optimisation algorithm, in this case a form of ‘Simplex’ [13], is then used to iteratively refine both the object location and orientation as well as the camera calibration parameters in order to maximise the quantity of image data consistent with the parameters. The covariance matrix is available for optimal combination of results with other calibration techniques. The validation process makes use of the information inherent in the full model, ignoring such factors as local stereo correspondence, concentrating instead on the refinement of the complete picture. This removes the assumptions inherent in the forward pass algorithms, which were necessary to allow the robust generation of the initial object location (Table 1). The remaining assumptions are that, after the forward pass, the model can be projected back onto the scene with sufficient accuracy to allow the CLV to find

Key	Original	Updated	Benefits
A	Canny [2] CONNECT	Canny [2] Bias Corrected Kalman Filter [11] Pollard [8]	- More reliable curve & line fitting
B	PMF GDB	Stretch Correlation [5]	Improved performance Temporal integration
C	GEOMstat	CLV	Independent of assumptions in Table 1
D	SMM	SMM + CLV	Robust statistics Top-down segmentation
E	Tsai [4]	Thacker [17]	Updatable on-line

Table 2: Summary of algorithm updates with key for Figs. 1 and 4.

the true location. This, in turn, assumes that the camera model is adequate. Put another way we now assume little more than the model is present in the scene and that we have a valid model of the camera optics. The final resulting statistic is a direct estimate of the quantity of edge data which is consistent with the model hypothesis and this is sufficient to corroborate the existence of the object at that location.

6 Experiments II

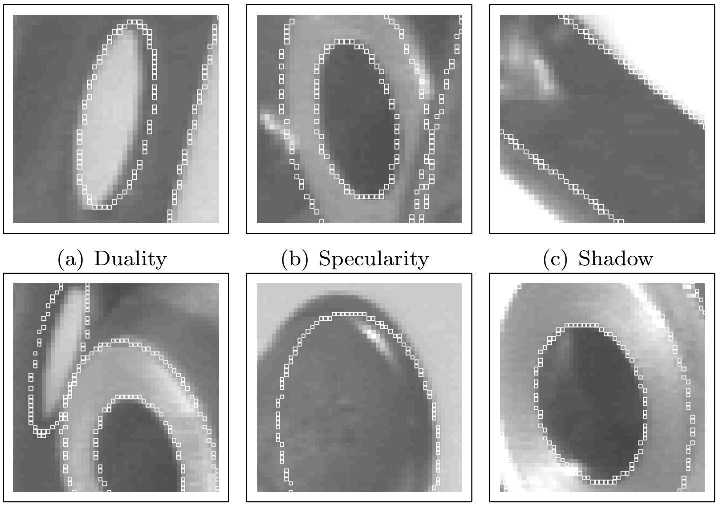
In this experiment the updated 3DMM with CLV was used to generate an initial hypothesis for the model location in both stereo images. An edge probability image was generated for the data. Matched features were projected onto these images and both the position of the object and the camera parameters were iteratively adjusted to give the best interpretation of the data. In comparison to the data shown in Fig. 3 it can be seen that the detail regions have provided consistent interpretations for the position of each 3D feature in the scene and have achieved agreement to within a pixel. In many cases it is obvious that the projected features do not locate onto edge data. This occurs when the feature is self occluded or view direction and illumination conspire to prevent the detection of a definite edge, i.e. the data isn't present. Non-the-less the combined requirement of each feature having to be maximally consistent with the projected model seems to be enough for the consistent model features to be selected in preference to those which are an artefact of viewpoint or illumination. The most significant change in the camera parameters adjusted by the CLV was a change of 5% in the aspect ratio of the cameras. Agreement between the model and the scene cannot be achieved without allowing this change. Modifying this parameter and re-running the stereo processing was not enough to eliminate gross errors in estimated model location. This suggests that the errors in the SMM are mainly due to inaccurate segmentation and statistical modelling of errors rather than simply an error in calibration. Thus the results achieved with the updated 3DMM could not be reached without the CLV, even if the correct camera calibration is used.

7 Discussion and Conclusions

A robust technique for accurately validating and locating a known object in a cluttered scene has been presented.

We have found that in order for algorithms to work well the assumptions (particularly statistical) must be valid and this often limits the success of closed form solutions, despite mathematical elegance. Iterative robust approaches are often necessary to eliminate the majority of problems. We have found that use of the most easily interpreted data in order to bootstrap the remaining data is generally the most efficient and reliable approach. At all stages the ability to estimate the accuracy of the resulting data is crucially important so that correct account can be taken of information in later stages of the system. View based wireframe model matching approaches to machine vision problems are no longer regarded as generic solutions and receive less attention in the literature. However, the statistical problems of data interpretation such as model selection and verification are still just as valid for rigid models as they are in the more recent approaches.

The motivations for working in an environment such as TINA are just as valid as they were when the system was conceived. The complexity of the vision problem is such that it can only be tackled in a modular fashion, where the overall problem is broken down into smaller, more manageable tasks. Solutions to these problem modules,



(a) Duality (b) Specularity (c) Shadow

(d) Specularity (e) Self-Occlusion (f) Specularity

Figure 6: Typical performance of the CLV model matcher on details from Fig. 3.

often research areas in their own right, result in sets of algorithms. Unless all the work is done within a unified environment and the constraints imposed by system integration are placed on the algorithmic design from the start, these individual solutions may never be made to work together. Put another way, system building generates the criteria and the environment necessary to evaluate an algorithm.

References

[1] P Burt and B Julesz. Modifications of the classical notion of panum's fusional area. *Perception*, 9:671–682, 1980.

[2] J F Canny. A computational approach to edge detection. *IEEE PAMI*, 8:679–714, 1986.

[3] S Crossley, A J Lacey, N A Thacker, and N L Seed. Robust stereo via temporal consistency. In *Proc. of the BMVC*, pages 659–668, 1997.

[4] An efficient and accurate camera calibration technique for 3d machine vision. R y tsai. In *Proc. IEEE CVPR 86*, pages 364–374, 1986.

[5] R A Lane, N A Thacker, and N L Seed. Stretch correlation as a real-time alternative to feature based stereo matching algorithms. *Image and Vision Computing*, 12(4):203–212, 1994.

[6] R A Lane, N A Thacker, N L Seed, and P A Ivey. A stereo vision processor. In *Proceedings of the CICC*, pages 181–184. IEEE, 1995.

[7] S B Pollard, J E W Mayhew, and J P Frisby. PMF: A stereo correspondence algorithm using a disparity gradient limit. *Perception*, 14:449–470, 1985.

[8] S B Pollard, J Porrill, and J E W Mayhew. Recovering partial wireframe descriptions from stereo data. In *Proceedings of the BMVC*, pages 31–36, 1990.

[9] J Porill, S B Pollard, T Pridmore, J Bowen, J E W Mayhew, and J P Frisby. *TINA: A 3D Vision System for Pick and Place.* MIT Press, 1991.

[10] J Porrill. Notes on the role of the disparity gradient in stereo vision. AIVRU LAb Memo 009, University of Sheffield, 1985.

[11] J Porrill. Fitting ellipses and predicting confidence envelopes using a bias corrected kalman filter. In *Proceedings of the Alvey Vision Conference*, pages 175–185, 1989.

[12] J Porrill, S B Pollard, T Pridmore J Bowen, J E W Mayhew, and J P Frisby. TINA: A 3D vision system for pick and place. In *Proceedings of the Alvey Vision Conference*, pages 65–72, 1987.

[13] W H Press, B P Flannery, S A Teukolsky, and W T Vetterling. *Numerical Recipes in C*. Cambridge University Press, New York, 1991.

[14] T Pridmore, J Porrill, and J E W Mayhew. Production rules for grouping edge-based disparity data. In *Proceedings of the Alvey Vision Conference*, 1985.

[15] F P Sykes, S B Pollard, and J E W Mayhew. Hypothesis and verification in 3-d model matching. In *Proc. 5th Alvey Vision Conference*, pages 7–12, 1989.

[16] N A Thacker and P Courtney. Calibration of a 4 DOF stereo head. In *Proceedings of the BMVC*, pages 528–537, 1992.

[17] N A Thacker and J E W Mayhew. Optimal combination of stereo camera calibration from arbitrary stereo images. *Image and Vision Computing*, 9(1):27–32, 1991.